A Latent Variable Representation of Count Data Models to Accommodate Spatial and Temporal Dependence: Application to Predicting Crash Frequency at Intersections

Marisol Castro
The University of Texas at Austin
Dept of Civil, Architectural and Environmental Engineering
1 University Station C1761, Austin, TX 78712-0278
Tel: 512-471-4535, Fax: 512-475-8744
Email: mcastro@mail.utexas.edu

Rajesh Paleti
The University of Texas at Austin
Dept of Civil, Architectural and Environmental Engineering
1 University Station C1761, Austin, TX 78712-0278
Tel: 512-471-4535, Fax: 512-475-8744
Email: rajeshp@mail.utexas.edu

Chandra R. Bhat*
The University of Texas at Austin
Dept of Civil, Architectural and Environmental Engineering
1 University Station C1761, Austin, TX 78712-0278
Tel: 512-471-4535, Fax: 512-475-8744
Email: bhat@mail.utexas.edu

*corresponding author

ABSTRACT

This paper proposes a reformulation of count models as a special case of generalized ordered-response models in which a single latent continuous variable is partitioned into mutually exclusive intervals. Using this equivalent latent variable-based generalized ordered response framework for count data models, we are then able to gainfully and efficiently introduce temporal and spatial dependencies through the latent continuous variables. Our formulation also allows handling excess zeros in correlated count data, a phenomenon that is commonly found in practice. A composite marginal likelihood inference approach is used to estimate model parameters. The modeling framework is applied to predict crash frequency at urban intersections in Arlington, Texas. The sample is drawn from the Texas Department of Transportation (TxDOT) crash incident files between 2003 and 2009, resulting in 1,190 intersection-year observations. The results reveal the presence of intersection-specific time-invariant unobserved components influencing crash propensity and a spatial lag structure to characterize spatial dependence. Roadway configuration, approach roadway functional types, traffic control type, total daily entering traffic volumes and the split of volumes between approaches are all important variables in determining crash frequency at intersections.

Keywords: Count data; multivariate analysis; spatial econometrics; accident analysis; composite marginal likelihood; generalized ordered response.

1. INTRODUCTION

Count data models rest on the assumption of a discrete probability distribution for the count variables, followed by the parameterization of the mean of the discrete distribution as a function of explanatory variables. Several types of discrete probability distributions may be considered in modeling count data, though the workhorse discrete distributions are the Poisson and the negative binomial (NB) distributions. Given that there is no *a priori* reason for the mean and variance of the count variable in any practical context to be equal, the use of a NB distribution is an important empirical generalization over the Poisson distribution. However, the variance of the NB distribution is higher than its mean, so that the NB distribution is applicable for over-dispersed data but not under-dispersed data. A discrete distribution that allows under-dispersion is the binomial distribution, though the binomial distribution requires an estimate of the maximum possible value of the count as an input. A discrete distribution that allows both under-dispersion and over-dispersion is the logarithmic distribution, but its mean has a relatively complicated form that makes it cumbersome to use when relating a count outcome as a function of exogenous variables. In addition to the distributions identified above, several modifications and generalizations of the Poisson and negative binomial distributions may also be used to accommodate under- and over-dispersion situations, as demanded by, and conceptually and theoretically appropriate to, the empirical context under consideration. These include the familiar zero-inflated count models (in which two separate states are identified for the count generating process– one that corresponds to a "zero" state in which the expected value of counts is so close to zero as being indistinguishable from zero, and another "normal" state in which a typical count model operates; see, for example, Musio *et al.*, 2010) and hurdle-count models (in which a two-part decision rule is postulated, in which a binary outcome process of the count being below or above a hurdle is combined with a truncated discrete distribution for the count process being above the hurdle point; see, for example, Bethell *et al.*, 2010).

The field has long matured in the area of univariate count models, with the approaches discussed above and their many variants (see, for example, Malyshkina and Mannering, 2010) already extensively used for univariate count data. However, this has not been the case for correlated count data, especially for the case of general dependency structures for more than two correlated counts. For instance, one may consider a simple Poisson or negative binomial discrete distribution, and develop multivariate versions of these discrete distributions to accommodate

correlated counts (see Buck *et al.*, 2009 and Bermúdez and Karlis, 2011 for applications of these methods). These multivariate Poisson and negative binomial distributions have the advantage of a closed form, but they become cumbersome as the number of correlated counts increases and they also represent the undesirable property that they can only accommodate a positive correlation in the counts. Alternatively, one may use a mixing structure, in which one or more random terms are introduced in the parameterization of the mean (so that the mean is not only a function of exogenous variables, but also includes one or more random terms within the exponentiation). If the same error term enters in the means of multiple count variables, this generates correlation. The most common form of such a mixture is to include normally distributed terms within the exponentiated mean function, so that the probability of the multivariate counts then requires integration over these random terms. The advantage of this method is that it permits both positive and negative dependency between the counts. Most studies incorporating such mixing structures have used a one-factor approach or used a small number of factors if a hierarchical clustering pattern is desired (see Wang *et al.*, 2006). However, such simple factor approaches to mixing impose coarse and restrictive dependency patterns on the count variables. Besides, it becomes difficult in the mixing approach (relative to non-mixing approaches) to multivariate count modeling to accommodate excess zeros through the use of techniques such as the zero-inflation method (see Herriges *et al.*, 2008).

In the current paper, we propose a count modeling framework and inference approach that resolves the many challenges discussed above for correlated counts. The presentation and application is motivated in the multivariate context of accommodating flexible spatial and temporal dependency patterns for a single count outcome variable, though it is equally applicable to other multivariate contexts in which correlated counts may arise. Specifically, we show how any traditional count model can be reformulated as a special case of a generalized ordered response model in which a single latent continuous variable is partitioned into mutually exclusive intervals. This is an issue that is not well understood and recognized in the literature. Using our equivalent latent variable-based generalized ordered response framework for count data models, we are then able to gainfully introduce spatial dependencies (using a spatial structure on the latent continuous variables) and time-stationary and time-varying temporal correlation patterns (by means of an appropriate structure for the error term of the latent variable). The authors are not aware of any study in the past that has accommodated such a

spatial structure as well as a flexible temporal error correlation pattern within the framework of count data models or generalized ordered-response models. The spatial and temporal dependencies in the resulting multivariate count framework leads to an analytically intractable likelihood function. In the current paper, we show how a composite marginal likelihood inference approach may be used for estimation. The approach is easy to implement and is based on evaluating lower-dimensional marginal probability expressions that do not require simulation (see Bhat *et al.*, 2010a, Varin *et al.*, 2011).

The proposed framework and inference approach is applied to study crash frequency at urban intersections, with the purpose of identifying the factors that contribute to intersection related-crashes. The crash data used in the analysis is drawn from the Texas Department of Transportation crash incident files for the City of Arlington and includes yearly crash frequency information over a 7-year period (2003 to 2009). In this context, the frequency of crashes at a particular intersection (say intersection A) may be inter-linked with those at other intersections over space because of at least two reasons (1) Spatially observed factors such as roadway geometry features or traffic flow characteristics at neighboring intersections may have a "spillover" effect on crash frequency at intersection A (even after accounting for the roadway/flow characteristics at intersection A) and (2) Spatially unobserved factors such as zonal regulations, neighborhood design features, and neighborhood driving attitudes that are not available to the researcher can cause a correlation between crash occurrence at proximately located intersections. Similarly, it is likely that intersection-specific unobserved factors (such as perhaps pedestrian walkway continuity characteristics or curb radius attributes or other roadway geometry features) cause a time-stationary and time-varying correlation in the number of crashes at the same intersection over time. We accommodate all these spatial and temporal effects, as well as spatial heterogeneity effects (which correspond to varying effects of exogenous variables on crash frequency across different intersections due to unobserved spatial effects).

The rest of the paper is structured as follows. The next section presents the building blocks of our approach in terms of model formulation and inference. Section 3 presents the model structure and estimation procedure. Section 4 illustrates an application of the proposed model for analyzing crash counts at urban intersections. The fifth and final section offers concluding thoughts and directions for further research.

2. THE BUILDING BLOCKS

2.1. A Unifying Latent Variable Framework for Ordered-Response and Count Models

In this section, we will develop a latent variable framework that brings the ordered response and count data models together in a simple cross-sectional context (spatial and temporal dependencies will be added later).

Let q ($q = 1, 2, \ldots, Q$) be an index to represent the observation unit and let k ($k = 0, 1, 2, 3, \ldots, K$) be an index to represent the ordinal level k of an ordered-response variable. The equation system for the standard ordered response (OR) model is (see Zavoina and McKelvey, 1975, who first proposed the OR model in its current form):

$$y_q^* = \boldsymbol{\beta}' \boldsymbol{x}_q + \varepsilon_q, \; y_q = k \text{ if } \psi_{k-1} < y_q^* < \psi_k, \tag{1}$$

where y_q^* corresponds to the latent propensity underlying the observed ordered variable. $\boldsymbol{x}_q$ is an ($L \times 1$)-column vector of exogenous attributes (excluding a constant). $\boldsymbol{\beta}$ is a corresponding ($L \times 1$)-column vector of variable effects. The latent propensity y_q^* is mapped to the observed ordinal variable y_q by the thresholds $\boldsymbol{\psi}$ ($\psi_{-1} = -\infty$ and $\psi_K = \infty$) in the usual ordered-response fashion. It is important to note that the model structure requires the thresholds to be strictly ordered for the partitioning of the latent risk propensity measure into the observed ordinal categories (*i.e.*, $-\infty < \psi_0 < \psi_1 < \psi_2 < \ldots < \psi_{K-1} < \infty$). ε_q is an idiosyncratic random error term that impacts the latent propensity and it is assumed to be identically and independently standard normal distributed across individuals q.[1]

In the standard ordered probit (SORP) model of Equation (1), the thresholds $\boldsymbol{\psi}$ are assumed to be fixed across individuals, though this need not be the case (see Terza, 1985, Pudney and Shields, 2000, King *et al.*, 2004, Kapteyn *et al.*, 2007, Eluru *et al.*, 2008, and King, 2009; see also Greene and Hensher, 2010, Chapter 7 for a discussion). Following these earlier studies, consider that the thresholds are parameterized as a non-linear function of a set of

---

[1] The exclusion of a constant in the vector $\boldsymbol{x}_q$ of Equation (1) is an innocuous normalization as long as all the intermediate thresholds (ψ_0 through ψ_{K-1}) are left free for estimation. Similarly, the use of the standard normal distribution rather than a non-standard normal distribution for the error term is also an innocuous normalization (see Zavoina and McKelvey, 1975; Greene and Hensher, 2010). Note also that any other proper continuous error distribution may be assumed for the error terms, such as the logistic distribution or the extreme value distribution. However, for our purpose of incorporating spatial and temporal dependencies later, the normal distribution has substantial benefits from an estimation standpoint. So, we will retain the normal distribution in the presentation here.

variables z_q (which includes a constant), $\psi_{qk} = f_k(z_q)$ (we include the index q to subscript the thresholds now to indicate that they are functions of the vector z_q). The non-linear nature of the functional form should ensure that the thresholds satisfy the ordering condition (*i.e.*, $-\infty < \psi_{q0} < \psi_{q1} < \psi_{q2} < \psi_{q,K-1} < \infty$), and also allow identification for any variables that are common in x_q and z_q.[2] There are several plausible reasons provided in the ordered-response literature to motivate such varying thresholds across observation units, all of which originate in the realization that the set of thresholds represents a dimension to introduce additional heterogeneity over and beyond the heterogeneity already embedded in the latent variable y_q^*. For instance, the threshold heterogeneity may be due to a different triggering mechanism for the translation (mapping) of the latent underlying y_q^* propensity variable to observed data or different perceptions (across respondents) of response categories in a survey. Such generalized threshold models are referred to by different names based on their motivating origins, but we will refer to them in the current paper as generalized ordered-response probit (GORP) models.

Now, consider a specific form of the GORP model system as follows:

$$y_q^* = \boldsymbol{\beta}' x_q + \varepsilon_q, \; y_q = k \text{ if } \psi_{q,k-1} < y_q^* < \psi_{qk}, \tag{2}$$

with $\psi_{qk} = f_k(z_q) = \Phi^{-1}\left(e^{-\lambda_q}\sum_{l=0}^{k}\frac{\lambda_q^l}{l!}\right)+\alpha_k$, where $\lambda_q = e^{\gamma' z_q}$, $k \in \{0,1,2,...,\infty\}$.

In the above equation, Φ^{-1} is the inverse function of the univariate cumulative standard normal, $\psi_{q,-1} = -\infty$, and $\alpha_0 = 0$ (the restriction $\alpha_0 = 0$ is imposed for identification given the parameterization of the ψ_{qk} terms; additional restrictions on the α_k parameters will generally be needed for estimation, as discussed later). γ is a coefficient vector to be estimated. The model in Equation (2) can exactly reproduce the traditional count data model with a Poisson discrete distribution with mean λ_q. To see this, assume that $\alpha_k = 0 \; \forall k$ and $\boldsymbol{\beta} = 0$. With these restrictions, the GORP model of Equation (2) collapses as follows:

---

[2] As indicated by Greene and Hensher (2010), the use of functional form to achieve identification is sometimes viewed with skepticism. On the other hand, there is nothing in the underlying theory of ordered-response models that requires the use of linear-in-parameters thresholds. The only requirement of the theory is the ordering of the thresholds, which, it so happens, requires some form of non-linear transformation to incorporate observed heterogeneity in the thresholds beyond the observed heterogeneity in the latent variable.

$$y_q^* = \varepsilon_q, \ y_q = k \ \text{if} \ \psi_{q,k-1} < y_q^* < \psi_{q,k}, \tag{3a}$$

$$\text{with} \ \psi_{qk} = f_k(z_q) = \Phi^{-1}\left(e^{-\lambda_q} \sum_{l=0}^{k} \frac{\lambda_q^l}{l!} \right), \text{where} \ \lambda_q = e^{\gamma' z_q}, \ k \in \{0,1,2,...,\infty\}, \psi_{q,-1} = -\infty.$$

Then, the probability expression in the GORP model of Equation (2) may be written as:

$$P[y_q = k] = P\left[\Phi^{-1}\left(e^{-\lambda_q} \sum_{l=0}^{k-1} \frac{\lambda_q^l}{l!} \right) < y_q^* < \Phi^{-1}\left(e^{-\lambda_q} \sum_{l=0}^{k} \frac{\lambda_q^l}{l!} \right) \right]$$

$$= \Phi\left(\Phi^{-1}\left(e^{-\lambda_q} \sum_{l=0}^{k} \frac{\lambda_q^l}{l!} \right) \right) - \Phi\left(\Phi^{-1}\left(e^{-\lambda_q} \sum_{l=0}^{k-1} \frac{\lambda_q^l}{l!} \right) \right) = \frac{e^{-\lambda_q} \lambda_q^k}{k!} = \frac{e^{-e^{\gamma' z_q}} \left(e^{\gamma' z_q} \right)^k}{k!} \tag{3b}$$

Essentially, then, by choosing the functional form for the parameterization of ψ_{qk} as in Equation (2), we have shown that the Poisson count regression model is exactly equivalent to a restricted version of the GORP model. This is an important result that has not been well understood and made explicit in earlier literature.

Several insights may be obtained from the recasting of a Poisson count model as a restricted version of the GORP framework. <u>First</u>, the reason traditional count models do not need any upper bound count value is that the thresholds in the equivalent latent regression framework of the GORP model are only functions of observation unit-specific variables (see Equation (3a)). On the other hand, in the SORP structure of Equation (1), the thresholds need to be estimated for each level k $(k = 1, 2, ..., K - 1)$. This feature of the SORP model has often been invoked as a disadvantage of the model for analyzing count data (see Winkelmann, 2000; page 70), because it requires an upper bound value K to be specified (by lumping counts above the value K to the value K). While this is true for the SORP framework, it need not be the case for the GORP model of Equation (2). Specifically, one can simply impose the restriction that $\alpha_k = 0 \ \forall k$ (as in Equation (3a)), or incorporate additional flexibility to accommodate high or low probability masses for specific outcomes by estimating some of the α_k parameters in the threshold function of Equation (2). All that needs to be done is to identify a value K above which α_k is held fixed at α_K; that is, $\alpha_k = \alpha_K$ for all $k > K$.[3] The analyst can empirically test different values of K and

---

[3] Note that K should be such that there are observed counts available in the sample for all consecutive count values from 0 to K.

compare data fit to determine the optimal value of K to add flexibility over the traditional Poisson count specification (that constrains all α_k parameters to zero). With such a specification of the threshold values, the GORP model in Equation (2) is a flexible Poisson count model and can predict the probability of an arbitrary count. Second, any discrete distribution-based count model may be used as the basis in the GORP model, even though the Poisson model has been used in Equation (2). The only requirement is that the thresholds be defined as:

$$\psi_{qk} = f_k(z_q) = \Phi^{-1}\left(\sum_{l=0}^{k}\left[P(y_q = k) \,|\, z_q\right]\right) + \alpha_k, \quad \text{where} \quad P(y_q = k) \,|\, z_q \quad \text{is the discrete probability}$$

distribution for outcome k. Third, the traditional count data approach to incorporate spatial dependency or temporal dependency or random coefficients involves a mixing structure in the threshold part of the latent framework. For instance, a general spatial dependency structure between the counts across the Q observation units may be introduced via a normal random error component η_q into the threshold parametrization as $\lambda_q = e^{\gamma' z_q + \eta_q}$. Collecting all the η_q terms into a single ($Q \times 1$)-vector $\boldsymbol{\eta} = (\eta_1, \eta_2, \eta_3, ..., \eta_Q)$, one can specify a multivariate normal distribution for $\boldsymbol{\eta}$; $\boldsymbol{\eta} \sim N_Q(0, \boldsymbol{\Omega})$. The resulting unconditional distribution for the multivariate probability of the observed outcome vector $\boldsymbol{m} = (m_1, m_2, m_3, ..., m_Q)$ is then obtained through the evaluation of the Q-dimensional integral:

$$P[\boldsymbol{m}] = \int_{\eta} \prod_{q=1}^{Q}\left[P(y_q = m_q) \,|\, \eta_q\right]\phi_Q(\boldsymbol{\eta})d\boldsymbol{\eta}, \tag{4}$$

where ϕ_Q is the Q-dimensional multivariate normal density function with mean 0 and covariance matrix $\boldsymbol{\Omega}$. The covariance matrix may be parameterized using a distance-based or similar spatial dependency structure (see Ver Hoef and Jansen, 2007 and Aguero-Valverde and Jovanis, 2010), but the model requires a Q-dimensional integration, which is all but impractical in most contexts even using recent simulation advances. If general temporal correlation is also specified for the case of Q observations units being observed for T periods each (in general, the number of time observations can differ between observational units, but we maintain the same number of time observations for ease in presentation here), the dimensionality for the probability expression grows further to QT. If random coefficients are additionally specified for the γ vector embedded in the threshold parameterization, the dimensionality is $QT+H$, where H is the number of random

coefficients in the $\boldsymbol{\gamma}$ vector. The estimation of models with such high dimensionality is next to infeasible with traditional simulation methods, particularly because of the highly non-linear fashion in which λ_q appears in the probability expression of Equation (3a). On the other hand, the recasting of the count data model in the latent variable framework opens up a new way to generate spatial and temporal dependencies in the count outcomes based on incorporating these dependencies in the specification of the underlying latent variables y_q^* rather than in the thresholds. Doing so allows for a more "linear" introduction of the dependencies and, as we will show later, is the key to being able to estimate flexibly correlated count data models.

The statistical benefits of using the GORL framework of Equation (2) for count data systems should be clear from the above discussion. However, the framework may also be motivated from an intuitive standpoint for count data in a manner similar to that for ordinal data. In the empirical context of crash counts at intersections, for example, one interpretation would be that there is a latent "long-term" (and constant over a certain time period) crash propensity y_q^* associated with intersection q that is a linear function of a set of intersection-related attributes $\boldsymbol{x}_q$. On the other hand, there may be some specific intersection characteristics (embedded in $\boldsymbol{z}_q$) that may dictate the likelihood of a crash occurring at any given *instant of time* for a given long-term crash propensity y_q^* (there may be common elements in $\boldsymbol{x}_q$ and $\boldsymbol{z}_q$). Thus, two intersections may have the same latent long-term crash propensity y_q^*, but may show quite different observed number of crashes over a certain time period because of different y_q^* - to - y_q mappings through the cut points (y_q is the observed count variable). From a latent variable framework perspective of Equation (3a), this is the view implicitly maintained by traditional count models. However, the traditional count models assume that the expected value of y_q^* is zero and constant across intersections (*i.e.*, in the notation of Equation (2), $y_q^* = \varepsilon_q$; that is, traditional count models assume that, up to a standard normal random term, all intersections have the same "long-term" zero value for the latent crash propensity). The GORP model relaxes this restriction, by allowing y_q^* to be linearly related to a vector of variables $\boldsymbol{x}_q$. Further, as will be clear later, our implicit assumption in extending the GORP framework to accommodate spatial dependency in counts is that it is the "long-term" latent crash propensity y_q^* that is responsible

for the spatial lag ("spillover") effects and the spatial correlation effects, not the elements that affect the "instantaneous" translation of the propensity to whether or not a crash occurs at any given time (and, therefore, not the threshold elements that affect the mapping of the latent propensity to the observed count outcome). Our expectation is that factors such as intersection traffic volumes, traffic control type and signal coordination, driveways between intersections, and roadway alignment are likely to affect "long-term" latent crash propensity at intersections and perhaps also the thresholds. These elements may also have a bearing on the "spillover" effects at other intersections, since they are likely to affect multiple intersections (in fact, these factors are important components considered in access management and roadway geometry strategies to improve safety at multiple intersections; see Chin and Quddus, 2003, and Huang and Chin, 2010). On the other hand, we postulate that there may be some specific intersection characteristics such as approach roadway types and curb radii at the intersection that will load more on the thresholds that affect the translation of the crash propensity to crash outcomes. Being intersection-specific, they also do not affect spatial spillover effects or spatial unobserved correlation effects. In terms of incorporating temporal dependency in counts from the same intersection, our formulation will retain the same variables (across time) in the latent propensity y_q^* and the thresholds. To the extent that many observed variables will either remain the same over time or be closely dependent on their earlier states, this will naturally generate temporal dependency in counts due to observed exogenous variables. We also expect that there will be time-invariant unobserved intersection-related factors affecting the long-term crash propensity y_q^*, as well as time-varying dependence in the effects of these unobserved factors based on temporal proximity. This is accommodated through an appropriate temporal error components specification for the long-term propensity. But we do not accommodate such unobserved temporal dependency effects in the thresholds, partly to avoid the highly non-linear random error component formulations that arise otherwise (as discussed earlier in this section in the context of traditional count models) and also because the thresholds represent "instantaneous" translation effects that we believe may not have strong temporal dependencies.

To summarize, the GORL framework represents a generalization of the traditional count data model, has the ability to retain all the desirable traits of count models and relax constraints imposed by count models, leads to a much simpler modeling structure when flexible spatial and

temporal dependencies are to be accommodated, and may also be justified from an intuitive/conceptual standpoint.

2.2. The Composite Marginal Likelihood Approach

The composite marginal likelihood (CML) estimation approach is a simple approach that can be used when the full likelihood function is near impossible or plain infeasible to evaluate due to underlying complex dependencies, as is the case of correlated count or ordered-response models of very high dimensionality discussed in the previous section. In this paper, we propose the use of the CML approach of estimation for count models within the GORP framework. The CML approach has been proposed for and applied to various forms of multivariate ordered-response model systems (see Varin and Czado, 2010, Bhat *et al.*, 2010a,b), but not for the kind of spatial and temporal dependency structures employed in this paper.

The CML approach, which belongs to the more general class of composite likelihood function approaches (see Lindsay, 1988), may be explained in a simple manner as follows. Assume that there are Q observation units and T time periods, and that data on each of the QT observation periods originates from a parametric underlying latent model based on a ($QT \times 1$) vector random variable y^* with density function $f(y^*, \theta)$, where θ is an unknown $\tilde{K}$-dimensional parameter vector. Based on this joint density, the appropriate likelihood function of the observed count outcomes m (*i.e.*, $y = m$, where m is a vector of dimension $QT \times 1$) may be written as $L(\theta, m)$. Suppose that this likelihood function is difficult or even infeasible to evaluate, but that evaluating the likelihood functions of subsets of the data characterized by the observable events ($A_1, A_2, ..., A_{\tilde{E}}$) are feasible and/or computationally expedient (for example, in the intersection accident count empirical application of the current paper, the observable events could correspond to the observations of accident count outcomes at a pair of intersections during the same year, or to the observations of a pair of accident count outcomes at the same intersection for two years). Let each event A_e be associated with a likelihood object $L_e(\theta, m \in A_e)$, which is based on a lower-dimensional marginal joint density function corresponding to the original high-dimensional joint density of y^*. Then, the general form of the composite marginal likelihood function is as follows:

$$L_{CML}(\boldsymbol{\theta}, \boldsymbol{m}) = \prod_{e=1}^{\widetilde{E}} \left[L_e(\boldsymbol{\theta}, \boldsymbol{m} \in A_e) \right] \tag{5}$$

The CML estimator is the one that maximizes the above function (or equivalently, its logarithmic transformation).

Almost all earlier research efforts employing the CML technique have used the pairwise approach in which the observed events A_e correspond to a pair of observations from the $(QT \times 1)$ vector $\boldsymbol{m}$. These earlier studies include Apanasovich *et al.*, (2008), Varin and Vidoni (2009), Engle *et al.* (2007), Bhat *et al.* (2010a), and Bhat and Sener (2009). Alternatively, the analyst can also consider larger subsets of observations, such as triplets or quadruplets or even higher dimensional subsets (see Engler *et al.*, 2006 and Caragea and Smith, 2007). However, it is generally agreed that the pairwise approach is a good balance between statistical and computational efficiency. The properties of the general CML estimator may be derived using the theory of estimating equations (see Cox and Reid, 2004, Yi *et al.*, 2011). Specifically, under usual regularity assumptions (Molenberghs and Verbeke, 2005, page 191, Xu and Reid, 2011), the CML estimator of $\boldsymbol{\theta}$ is consistent and asymptotically normal distributed with asymptotic mean $\boldsymbol{\theta}$ and covariance matrix given by the inverse of Godambe's (1960) sandwich information matrix (see Zhao and Joe, 2005):

$$V_{CML}(\hat{\boldsymbol{\theta}}) = [\boldsymbol{G}(\boldsymbol{\theta})]^{-1} = [\boldsymbol{H}(\boldsymbol{\theta})]^{-1} \boldsymbol{J}(\boldsymbol{\theta}) [\boldsymbol{H}(\boldsymbol{\theta})]^{-1}, \tag{6}$$

where $\boldsymbol{H}(\boldsymbol{\theta}) = E\left[-\dfrac{\partial^2 \log L_{CML}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right]$ and $\boldsymbol{J}(\boldsymbol{\theta}) = E\left[\left(\dfrac{\partial \log L_{CML}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right) \left(\dfrac{\partial \log L_{CML}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} \right) \right]$.

3. THE MODEL

In the current section, we introduce spatial dependence through the latent crash propensity variable, using the GORP framework established in Section 2.1 for count data modeling. We consider the spatial lag error structure in the paper, which allows spatial dependence through both spatial "spillover" effects caused by observed exogenous variables at one location impacting the dependent variable of interest at another location as well as spatial error correlation effects caused by unobserved attributes at one location impacting the dependent variable at another location.

Several studies in the past decade have considered the spatial lag error structure for binary choice models (for example, see Fleming, 2004, Franzese and Hays, 2008, Franzese *et al.*, 2010, and LeSage and Pace, 2009 for good reviews). The two dominant techniques, both based on simulation methods, for the estimation of the spatial lag model are the frequentist recursive importance sampling (RIS) estimator (which is a generalization of the more familiar Geweke-Hajivassiliou-Keane or GHK simulator; see Beron *et al.*, 2003 and Beron and Vijverberg, 2004) and the Bayesian Markov Chain Monte Carlo (MCMC)-based estimator (see LeSage and Pace, 2009). However, both of these methods are confronted with multi-dimensional normal integration. The RIS and MCMC estimators are cumbersome and sometimes even infeasible to implement in typical empirical contexts, because of the high dimensional integration needed (1190 dimensions in the current empirical context).[4]

The next section presents the model formulation for both the spatial lag and spatial error structures, while Section 3.2 discusses model estimation.

3.1. The Spatial Lag Count Model with Temporal Dependence

As earlier, let q be an index for observation units ($q = 1, 2, \ldots, Q$). We now add an index t for time period ($t = 1, 2, \ldots, T$).[5] Let the observed count for unit q at the tth period be m_{qt} (m_{qt} may take any non-negative integer value). Consider the following structure for the latent underlying and continuous variable y_{qt}^* in the GORP representation for count models:

$$y_{qt}^* = \delta \sum_{q'=1}^{Q} w_{qq'} y_{q't}^* + \boldsymbol{\beta_q'} \boldsymbol{x_{qt}} + \varepsilon_{qt}, \quad y_{qt} = m_{qt} \text{ if } \psi_{q,m_{qt}-1,t} < y_{qt}^* < \psi_{q,m_{qt},t}, \tag{7}$$

where $w_{qq'}$ is the usual distance-based spatial weight corresponding to units q and q' (with $w_{qq} = 0$ and $\sum_{q'} w_{qq'} = 1$) for each (and all) q, and δ $(0 < \delta < 1)$ is the spatial autoregressive

---

[4] Many studies attempt to side-step the high dimensional problem by clustering observation units into "regions", and then considering a spatial error dependency over the regions rather than the observational units (see Smith and LeSage, 2004 and Phaneuf and Palmquist, 2003).

[5] We assume here that the number of periods of observation is the same across individual units, as is the case in the empirical analysis of the current paper. Extension to the case of different numbers of panel observations across units does not pose any substantial challenges in formulation or estimation. The only minor difference is that a weight needs to be placed for each unit in the CML estimation approach (see Kuk and Nott, 2000, Joe and Lee, 2009, and Bhat, 2011).

parameter, x_{qt} is a ($L\times 1$)-vector of exogenous variables (including a constant now to accommodate time-stationary random effects through a random coefficient on this constant) and $\boldsymbol{\beta}_q$ is an observation unit-specific ($L\times 1$)-vector of coefficients assumed to be a realization from a multivariate normal distribution with mean vector $\boldsymbol{b}$ and covariance $\boldsymbol{\Omega} = \boldsymbol{LL'}$. It is not necessary that all elements of $\boldsymbol{\beta}_q$ be random; that is, the analyst may specify fixed coefficients on some exogenous variables in the model, though it will be convenient in presentation to assume that all elements of $\boldsymbol{\beta}_q$ are random. Also, note that the element of $\boldsymbol{b}$ corresponding to the constant is fixed to zero for identification. This is similar to Equation (1), where we did not include a constant in the vector x_q given the parameterization of the thresholds; however, here we will find it convenient for presentation to absorb the constant in the x_{qt} vector and fix its mean coefficient to zero. The variance of the coefficient on the constant captures time-invariant dependence in the outcomes measured on the same decision unit q. For later use, we will write $\boldsymbol{\beta}_q = \boldsymbol{b} + \widetilde{\boldsymbol{\beta}}_q$, where $\widetilde{\boldsymbol{\beta}}_q \sim MVN_L(0,\boldsymbol{\Omega})$ (MVN_L represents the multivariate normal distribution of dimension L). ε_{qt} in Equation (7) is a standard normal error term uncorrelated with $\widetilde{\boldsymbol{\beta}}_q$ and uncorrelated across observation units q, but with a first-order autoregressive temporal dependence structure for the error terms of the same observation unit q ($\mathrm{corr}(\varepsilon_{qt}, \varepsilon_{qs}) = \rho^{|t-s|} \, (0 < \rho < 1)$). The thresholds in Equation (7) take the form discussed earlier:

$$\psi_{q,m_{qt},t} = \Phi^{-1}\left(e^{-\lambda_{qt}} \sum_{l=0}^{m_{qt}} \frac{\lambda_{qt}^l}{l!} \right) + \alpha_{m_{qt}}, \quad \lambda_{qt} = e^{\gamma' z_{qt}}, \alpha_0 = 0, \alpha_{m_{qt}} = \alpha_K \text{ if } m_{qt} > K, \tag{8}$$

where z_{qt} is a vector of exogenous variables (including a constant) associated with observation unit q at the tth time period, $\boldsymbol{\gamma}$ is a corresponding coefficient vector to be estimated, and K is a pre-defined count level as discussed in Section 2.1. Equations (7) and (8) represent a count framework that accommodates a spatial lag structure, time-invariant and time-varying dependencies, as well as random coefficients.

To write the equation system in (7) compactly, we next define several vectors and matrices. Let $\boldsymbol{y}_t^* = (y_{1t}^*, y_{2t}^*, y_{3t}^*, ..., y_{Qt}^*)'$ and $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, \varepsilon_{2t}, \varepsilon_{3t}, ..., \varepsilon_{Qt})'$ be ($Q\times 1$) vectors, and let $\boldsymbol{x}_t = (\boldsymbol{x}_{1t}, \boldsymbol{x}_{2t}, \boldsymbol{x}_{3t}, ..., \boldsymbol{x}_{Qt})'$ be a ($Q\times L$) matrix of exogenous variables for all Q units for the tth

13

time period. Next, let $y^* = [(y_1^*)', (y_2^*)', (y_3^*)', ..., (y_T^*)']'$ ($QT \times 1$ vector), $\varepsilon = (\varepsilon_1', \varepsilon_2', \varepsilon_3', ..., \varepsilon_T',)'$ ($QT \times 1$ vector), $x = (x_1', x_2', x_3', ..., x_T')'$ ($QT \times L$ matrix), and $\widetilde{\beta} = (\widetilde{\beta}_1', \widetilde{\beta}_2', \widetilde{\beta}_3', ..., \widetilde{\beta}_Q')'$ ($QL \times 1$ vector). Define y as the outcome corresponding to the latent vector y^*. Define $\widetilde{x}_t$ as a $Q \times (L \times Q)$ block-diagonal matrix with each block-diagonal of size ($1 \times L$) being occupied by the vector x_{qt}' (q = 1, 2, ..., Q), and let $\widetilde{x} = (\widetilde{x}_1', \widetilde{x}_2', \widetilde{x}_3', ..., \widetilde{x}_T')'$ ($QT \times QL$ matrix). Collect all the weights $w_{qq'}$ into a spatial weight matrix W. With these definitions, the latent regression part of Equation (7) may be re-written as:

$$y^* = \delta(I_T \otimes W)y^* + xb + \widetilde{x}\widetilde{\beta} + \varepsilon ,\tag{9}$$

where I_T is an identity matrix of size T. After further matrix manipulation to write y^* in reduced form, we obtain:

$$y^* = Sxb + S\widetilde{x}\widetilde{\beta} + S\varepsilon, \text{ where } S = [I_{QT} - \delta(I_T \otimes W)]^{-1} = I_T \otimes [(I_Q - \delta W)^{-1}].\tag{10}$$

The expected value and variance of y^* may be obtained from the above equation after developing the covariance matrix for the error vector ε. To do so, note that the error vector ε is distributed multivariate normal with a mean vector of zero and a temporal autoregressive covariance matrix $\Lambda \otimes I_Q$ (of size $QT \times QT$), where Λ is given by:

$$\Lambda = \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 & \cdots & \rho^{T-1} \\ \rho & 1 & \rho & \rho^2 & \cdots & \rho^{T-2} \\ \rho^2 & \rho & 1 & \rho & \cdots & \rho^{T-3} \\ \vdots & \vdots & \vdots & 1 & \cdots & \vdots \\ \rho^{T-1} & \rho^{T-2} & \rho^{T-3} & \rho^{T-4} & \cdots & 1 \end{bmatrix}\tag{11}$$

Then, we obtain $y^* \sim MVN_{QT}(B, \Sigma)$, where

$$B = Sxb \text{ and } \Sigma = S[\widetilde{x}(I_Q \otimes \Omega)\widetilde{x}' + \Lambda \otimes I_Q] S'\tag{12}$$

The structure of the covariance matrix above generates a dependency across all QT count outcomes. Specifically, the S matrix is responsible for the spatial dependence across observational units for each time period, while the sandwich matrix in Equation (12) generates the temporal dependence for each observation unit across the T time periods (due to the time-

invariant temporal dependence effects of the random coefficients, and the time-varying autoregressive temporal dependence effect). At the same time, the random coefficients allow unobserved heterogeneity (across observational units) in the effects of exogenous variables on the latent crash propensity.

3.2. Model Estimation

The parameter vector to be estimated in the spatial lag model is $\theta = (\boldsymbol{b}', \overline{\boldsymbol{\Omega}}', \rho, \delta, \boldsymbol{\gamma}', \boldsymbol{\alpha}')'$, where $\overline{\boldsymbol{\Omega}}$ is a column vector obtained by vertically stacking the upper triangle elements of the matrix $\boldsymbol{\Omega}$ and $\boldsymbol{\alpha}$ is another column vector obtained by vertically stacking the α_k parameters. Several restrictive models are obtained from the model developed here. If $\rho = 0$, this indicates lack of time-varying temporal correlation in the counts. If $\delta = 0$, the result is a non-spatial model. If the elements of $\overline{\boldsymbol{\Omega}}$ are zero, the indication is the lack of time-invariant temporal effects as well as unobserved heterogeneity across observational units. If the elements of $\overline{\boldsymbol{\Omega}}$ corresponding to the non-diagonal elements of $\boldsymbol{\Omega}$ are zero, but not the diagonal elements, it represents the case of the presence of time-invariant and unobserved heterogeneity effects, but without correlation between these effects. If $\rho = 0$, $\delta = 0$, and all elements of $\overline{\boldsymbol{\Omega}}$ are zero, this corresponds to a flexible count model (recast in the GORP framework) with no spatial, temporal, and unobserved heterogeneity effects. Further, if all elements of $\boldsymbol{\gamma}$ are zero (except the one on the constant) and the analyst specifies an upper bound K for the counts (both in the estimation sample and for prediction purposes), the result is a standard ordered-response model. Finally, if $\rho = 0$, $\delta = 0$, and all elements of $\overline{\boldsymbol{\Omega}}$, $\boldsymbol{b}$ and $\boldsymbol{\alpha}$ are simultaneously zero, the result is a traditional count model.

The likelihood function for the model is:

$$L(\boldsymbol{\theta}) = P(\boldsymbol{y} = \boldsymbol{m}) = \int_{D_{y^*}} \phi_{QT}(\boldsymbol{y}^* \mid \boldsymbol{b}, \boldsymbol{\Sigma}) d\boldsymbol{y}^*, \tag{13}$$

where $D_{y^*} = \{\boldsymbol{y}^* : \psi_{(q,m_{qt}-1,t)} < y_{qt}^* < \psi_{q,m_{qt},t}, \ \forall q = 1, 2, ..., Q, \ t = 1, 2, ..., T\}$ and $\phi_{QT}(.)$ is the multivariate normal density function of dimension QT. $\boldsymbol{m}$ is a $QT \times 1$-vector of observed count outcomes as follows: $\boldsymbol{m} = (m_{11}, m_{21}, m_{31}, ..., m_{Q1}, m_{12}, m_{22}, m_{32}, ..., m_{Q2}, ..., m_{1T}, m_{2T}, m_{3T}, ..., m_{QT})'$. The integration domain D_{y^*} is simply the multivariate region of the elements of the $\boldsymbol{y}^*$ vector

determined by the observed vector of count outcomes. The dimensionality of the rectangular integral in the likelihood function is QT. The numerical evaluation of this integral can become problematic, even for moderately sized Q and T. The use of numerical simulation techniques based on a maximum simulated likelihood (MSL) (see Greene, 2005) or a Bayesian inference approach, even if feasible, can lead to convergence problems during estimation (Bhat *et al.*, 2010a; Müller and Czado, 2005). The alternative is to use the composite marginal likelihood (CML) approach, as discussed in Section 2.2. In the current study, we use the pairwise composite marginal likelihood method based on the product of the likelihood contributions from pairs of observation units across time periods. To write this function, define two threshold vectors of size $QT \times 1$ as follows:

$$\boldsymbol{\tau} = (\psi_{1,m_{11}-1,1}, \psi_{2,m_{21}-1,1}, \ldots, \psi_{Q,m_{Q1}-1,1}, \psi_{1,m_{12}-1,2}, \psi_{2,m_{22}-1,2}, \ldots, \psi_{Q,m_{Q2}-1,2}, \ldots \psi_{1,m_{1T}-1,T}, \psi_{2,m_{2T}-1,T}, \ldots, \psi_{Q,m_{QT}-1,T})',$$

$$\boldsymbol{\vartheta} = (\psi_{1,m_{11},1}, \psi_{2,m_{21},1}, \ldots, \psi_{Q,m_{Q1},1}, \psi_{1,m_{12},2}, \psi_{2,m_{22},2}, \ldots, \psi_{Q,m_{Q2},2}, \ldots \psi_{1,m_{1T},T}, \psi_{2,m_{2T},T}, \ldots, \psi_{Q,m_{QT},T})'.$$

Let g be an index that can takes the values from 1 to QT. Then,

$$
\begin{aligned}
L_{CML}(\boldsymbol{\theta}) &= \left(\prod_{g=1}^{QT-1} \prod_{g'=g+1}^{QT} P\big([\boldsymbol{y}]_g = [\boldsymbol{m}]_g, [\boldsymbol{y}]_{g'} = [\boldsymbol{m}]_{g'} \big) \right) \\
&= \left(\prod_{g=1}^{QT-1} \prod_{g'=g+1}^{QT} \begin{bmatrix} \Phi_2(\varphi_g, \varphi_{g'}, \nu_{gg'}) - \Phi_2(\varphi_g, \mu_{g'}, \nu_{gg'}) \\ -\Phi_2(\mu_g, \varphi_{g'}, \nu_{gg'}) + \Phi_2(\mu_g, \mu_{g'}, \nu_{gg'}) \end{bmatrix} \right),
\end{aligned}
\tag{14}
$$

where $\varphi_g = \dfrac{[\boldsymbol{\vartheta}]_g - [\boldsymbol{B}]_g}{\sqrt{[\boldsymbol{\Sigma}]_{gg}}}$, $\mu_g = \dfrac{[\boldsymbol{\tau}]_g - [\boldsymbol{B}]_g}{\sqrt{[\boldsymbol{\Sigma}]_{gg}}}$, $\nu_{gg'} = \dfrac{[\boldsymbol{\Sigma}]_{gg'}}{\sqrt{[\boldsymbol{\Sigma}]_{gg}}\sqrt{[\boldsymbol{\Sigma}]_{g'g'}}}$.

In the above expression, $[\boldsymbol{\vartheta}]_g$ represents the gth element of the column vector $\boldsymbol{\vartheta}$, and similarly for other vectors. $[\boldsymbol{\Sigma}]_{gg'}$ represents the gg^{th} element of the matrix $\boldsymbol{\Sigma}$. The CML estimator is obtained by maximizing the logarithm of the function in Equation (14). Note that, unlike other simulation-based estimators, the CML estimator entails only the computation of bivariate cumulative normal distribution functions, which are extremely quick to evaluate.

The pairwise marginal likelihood function of Equation (14) comprises $QT(QT-1)/2$ pairs of bivariate probability computations, which can itself become quite time consuming. Fortunately, in a spatial-temporal case where spatial dependency drops quickly with inter-observation distance, the pairs formed from the closest spatial observation units provide much more information than pairs from spatial units that are far away. In fact, as demonstrated by

Varin and Vidoni (2009), Bhat *et al.* (2010a), and Varin and Czado (2010) in different empirical contexts, retaining all pairs not only increases computational costs, but may also reduce estimator efficiency. We examine this issue by creating different distance bands and, for each specific distance band, considering only those count pairings in the CML function that are within the distance band. To do so, construct a $Q \times Q$ matrix $\widetilde{R}$ with its qth column filled with a $Q \times 1$ vector of zeros and ones as follows: if the observational unit q' is not within the specified threshold distance of unit q, the q'^{th} row has a value of zero; otherwise, the q'^{th} row has a value of one. By construction, the qth row of the qth column has a value of one. Also, let $R = \mathbf{1}_{T \times T} \otimes \widetilde{R}$, where $\mathbf{1}_{T \times T}$ is a $T \times T$-matrix of ones. Then, the CML function gets modified as follows:

$$L_{CML}(\boldsymbol{\theta}) = \left(\prod_{g=1}^{QT-1} \prod_{g'=g+1}^{QT} \left[P\left([\boldsymbol{y}]_g = [\boldsymbol{m}]_g, [\boldsymbol{y}]_{g'} = [\boldsymbol{m}]_{g'} \right) \right]^{[R]_{gg'}} \right)$$
$$= \left(\prod_{g=1}^{QT-1} \prod_{g'=g+1}^{QT} \left[\begin{array}{c} \Phi_2(\varphi_g, \varphi_{g'}, \boldsymbol{v}_{gg'}) - \Phi_2(\varphi_g, \boldsymbol{\mu}_{g'}, \boldsymbol{v}_{gg'}) \\ -\Phi_2(\boldsymbol{\mu}_g, \varphi_{g'}, \boldsymbol{v}_{gg'}) + \Phi_2(\boldsymbol{\mu}_g, \boldsymbol{\mu}_{g'}, \boldsymbol{v}_{gg'}) \end{array} \right]^{[R]_{gg'}} \right).$$
(15)

We develop the asymptotic variance matrix $V_{CML}(\hat{\boldsymbol{\theta}})$ of Equation (6) (as discussed next) for each distance band and select the distance band that minimizes the total variance across all parameters as given by $tr[V_{CML}(\hat{\boldsymbol{\theta}})]$, where $tr[A]$ denotes the trace of the matrix A.

The asymptotic variance expression is given by the sandwich estimator (see Section 2.2). The "bread" matrix $H(\boldsymbol{\theta})$ of Equation (6) can be estimated in a straightforward manner using the Hessian of the negative of $\log L_{CML}(\boldsymbol{\theta})$, evaluated at the CML estimate $\hat{\boldsymbol{\theta}}$:

$$\hat{\boldsymbol{H}}(\hat{\boldsymbol{\theta}}) = -\left[\sum_{g=1}^{QT-1} \sum_{g'=g+1}^{QT} \frac{\partial^2 \log L_{CML,gg'}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right]_{\hat{\boldsymbol{\theta}}},$$
(16)

where $L_{CML,gg'}(\boldsymbol{\theta}) = \left[P\left([\boldsymbol{y}]_g = [\boldsymbol{m}]_g, [\boldsymbol{y}]_{g'} = [\boldsymbol{m}]_{g'} \right) \right]^{R_{gg'}} | \boldsymbol{\theta}$.

However, the estimation of the "vegetable" matrix $J(\boldsymbol{\theta})$ is not straightforward because of the underlying spatial and temporal dependence among counts. But, because the spatial dependence pattern implied by the spatial lag structure fades with distance, one can use the

windows re-sampling procedure of Heagerty and Lumley (2000) to estimate $\boldsymbol{J}(\boldsymbol{\theta})$. This procedure entails the construction of suitable overlapping subgroups of the count data that may be viewed as independent replicated observations. While there are several methods to do so, we use the method proposed by Bhat (2011). His approach is to overlay the spatial region under consideration with a square grid providing a total of D internal and external nodes. Then, select the observational unit closest to each of the D grid nodes to obtain D observational units from the original Q observational units ($d = 1, 2, 3, \ldots, D$). Let $\widetilde{C}$ be a $Q \times D$ matrix with its dth column filled with a $Q \times 1$ vector of zeros and ones, with a zero value in the q'^{th} row ($q' = 1, 2, \ldots, Q$) if the observational unit q' is not within the specified threshold distance of unit d, and a one otherwise (by construction, $\widetilde{C}_{q'd} = 1$ if $q' = d$). Also, let $\boldsymbol{C} = \mathbf{1}_T \otimes \widetilde{C}$, where $\mathbf{1}_T$ is a $T \times 1$-matrix of ones. Then, the columns of $\boldsymbol{C}$ provide pseudo-independent sets of observational units.[6] Let the score matrix corresponding to the pairings in column d of matrix $\boldsymbol{C}$ be $\boldsymbol{s}_{CML,d}(\boldsymbol{\theta})$. Also, let N_d be the sum of the dth column of $\boldsymbol{C}$, and let $\widetilde{W} = \sum_{g=1}^{QT-1} \sum_{g'=g+1}^{QT} [\boldsymbol{R}]_{gg'}$. Then, the $\boldsymbol{J}(\boldsymbol{\theta})$ matrix may be empirically estimated as:

$$\boldsymbol{J}(\hat{\boldsymbol{\theta}}) = \frac{\widetilde{W}}{D} \left[\sum_{d=1}^{D} \left[\frac{1}{N_d} \left([\boldsymbol{s}_{CML,d}(\boldsymbol{\theta})][\boldsymbol{s}_{CML,d}(\boldsymbol{\theta})]' \right)_{\hat{\boldsymbol{\theta}}} \right] \right]. \tag{17}$$

One final important issue that we have not discussed thus far is how to ensure the positive definiteness of the matrix $\boldsymbol{\Omega}$. Once this is ensured, the positive definiteness of $\boldsymbol{\Sigma}$ is ensured as long as $0 < \delta < 1$ and $0 < \rho < 1$. In our estimation, the positive-definiteness of $\boldsymbol{\Omega}$ is guaranteed by writing the logarithm of the pairwise-likelihood in terms of the Cholesky-decomposed elements of $\boldsymbol{\Omega}$ and maximizing with respect to these elements of the Cholesky factor. Essentially, this procedure entails passing the Cholesky elements as parameters to the optimization routine, constructing the $\boldsymbol{\Omega}$ matrix internal to the optimization routine, then computing $\boldsymbol{\Sigma}$, and finally picking off the appropriate elements of the matrix for the pairwise

---

[6] As indicated by Bhat (2011), there needs to be a balance here between the number of sets of pairings D and the proximity of points. The smaller the value of D, the less proximal are the sets of observation units and more likely that the sets of observational pairings will be independent. However, at the same time, the value of D needs to be reasonable to obtain a good empirical estimate of $\boldsymbol{J}$, since this empirical estimate is based on averaging the cross-product of the score functions (computed at the convergent parameter values) across the D sets of observations.

likelihood components. To ensure the constraints on the autoregressive terms δ and ρ, we parameterize these terms as $\delta = 1/[1 + \exp(\widetilde{\delta})]$ and $\rho = 1/[1 + \exp(\widetilde{\rho})]$, respectively. Once estimated, the $\widetilde{\delta}$ and $\widetilde{\rho}$ estimates can be translated back to estimates of δ and ρ.

3.3. Model Selection

Procedures similar to those available with the maximum likelihood approach are also available for model selection with the CML approach (see Varin and Vidoni, 2009, Pace *et al.*, 2011 and Bhat, 2011). The statistical test for a single parameter may be pursued using the usual t-statistic. When the statistical test involves multiple parameters between two nested models, an appealing statistic, which is also similar to the likelihood ratio test in ordinary maximum likelihood estimation, is the adjusted composite likelihood ratio test (*ADCLRT*) statistic. Consider the null hypothesis $H_0 : \tau = \tau_0$ against $H_1 : \tau \neq \tau_0$, where τ is a subvector of $\boldsymbol{\theta}$ of dimension p. Let $\hat{\boldsymbol{\theta}}$ be the CML estimator of the unrestricted model (without the restriction imposed by the null hypothesis), and let $\hat{\boldsymbol{\theta}}_0$ be the CML estimator under the null hypothesis. Define $[\boldsymbol{G}_\tau(\boldsymbol{\theta})]^{-1}$ and $[\boldsymbol{H}_\tau(\boldsymbol{\theta})]^{-1}$ as the $p \times p$ submatrices of $[\boldsymbol{G}(\boldsymbol{\theta})]^{-1}$ and $[\boldsymbol{H}(\boldsymbol{\theta})]^{-1}$, respectively, which correspond to the vector τ. The following adjusted *CLRT* statistic, *ADCLRT*, may be considered to be asymptotically chi-squared distributed with p degrees of freedom:

$$ADCLRT = \frac{[\boldsymbol{S}_\tau(\boldsymbol{\theta})]'[\boldsymbol{H}_\tau(\boldsymbol{\theta})]^{-1}[\boldsymbol{G}_\tau(\boldsymbol{\theta})][\boldsymbol{H}_\tau(\boldsymbol{\theta})]^{-1}\boldsymbol{S}_\tau(\boldsymbol{\theta})}{[\boldsymbol{S}_\tau(\boldsymbol{\theta})]'[\boldsymbol{H}_\tau(\boldsymbol{\theta})]^{-1}\boldsymbol{S}_\tau(\boldsymbol{\theta})} \times CLRT, \tag{18}$$

where $CLRT = 2[\log L_{CML}(\hat{\boldsymbol{\theta}}) - \log L_{CML}(\hat{\boldsymbol{\theta}}_0)]$, $\boldsymbol{s}_\tau(\boldsymbol{\theta})$ is the $p \times 1$ sub-vector of $\boldsymbol{s}(\boldsymbol{\theta}) = \left(\dfrac{\partial \log L_{CML}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)$ corresponding to the vector τ, and all the matrices above are computed at $\hat{\boldsymbol{\theta}}_0$.

4. APPLICATION TO INTERSECTION ACCIDENT COUNTS

4.1. Background

Motorized vehicle travel is the principal means of personal transportation in the United States. Although providing mobility and accessibility to activities, motorized vehicle travel also carries

with it the risk of being involved in a roadway crash, leading to deaths, injuries and property damage. According to the National Highway Traffic Safety Administration (NHTSA), in 2009 alone, 33,808 people were killed and 2.2 million people injured in roadway crashes (NHTSA, 2010a). This translates to an average of 93 fatalities per day in motor vehicle crashes in 2009; that is, one fatality every 16 minutes on US roadways. In fact, motor vehicle crashes continue to be the leading cause of death for people aged 11 through 33 years of age (NHTSA, 2010a). Even if a roadway accident does not involve a fatality, accidents represent an enormous cost to society, including property and motor vehicle damage, productivity losses, medical and administrative expenses, mental trauma, pain, and increased insurance premiums. The U.S. National Safety Council (2009) has quantified the average cost to society per death as $1,290,000 and per nonfatal disabling injury as $68,100. The overall economic cost of roadway-related crashes is estimated to be over $190 billion per year.

Among all traffic accidents, intersection and intersection-related crashes make up about 40% of total crashes (NHTSA, 2010b). This is not surprising, because intersections generate conflicts of movement, are locations of stop-and-go traffic, and correspond to roadway locations with dense traffic. In the pool of serious intersection crashes (those involving one or more fatalities), 60% occur at urban intersections. Thus, understanding the causes of intersection related crashes in general, and in urban areas in particular, should be a priority for transportation and safety professionals in developing crash countermeasures.

Indeed, the study of crashes at intersection locations has received increasing attention in recent years. Several studies (see, for example, Haque *et al.*, 2010, Huang and Chin, 2010, Mitra, 2009, Wang and Abdel-Aty, 2006, Chin and Quddus, 2003, and Griebe, 2003), have examined the number of crashes occurring at an intersection as a function of intersection control characteristics, roadway design features, and traffic volumes. However, many of these studies assume that intersections are completely isolated entities, with no spatial dependence in the frequency of crashes between proximally located intersections. Some other studies consider spatial dependency, but in a rather coarse and restrictive form by assuming that the crashes at intersections within a certain geographic region or location type (such as primarily business, primary residential, primarily retail, and open county) or other spatial clusterings are correlated due to unobserved locational factors. But this approach assumes a constant level of correlation in crash risk across intersections within a spatial cluster, and no correlation between intersections in

20

different spatial clusters. Mitra (2009) considers a global spatial dependence configuration where the crashes at one intersection are related to all other intersections based on unobserved spatial proximity effects.[7] However, Mitra's approach does not accommodate a spatial lag structure, includes a single covariate (traffic flow), does not consider time-varying temporal effects or heterogeneity effects in the impact of variables, and uses a traditional Poisson count model with no accommodation of excess zero counts. The model in Mitra's study is estimated using Bayesian MCMC methods, which have their limitations as discussed earlier. More generally, the fundamental problem with past approaches is that the mixing of error terms in the log-risk formulation (to generate the needed spatial dependence covariance over a Poisson or negative binomial or related versions of these kernel count models) leads to a very cumbersome likelihood function. Simulation techniques (both MSL and Bayesian) are not suited to handle the estimation of such models unless restrictive assumptions are imposed on the nature of spatial dependence.

4.2. Data

The crash data used in the analysis is drawn from the Texas Department of Transportation (TxDOT) Crash Records Information System (CRIS) between 2003 and 2009. The CRIS compiles police and driver reports of crashes into multiple text files, including complete crash, vehicle, person, and weather-related details for each crash.[8] TxDOT overlays the crash locations from the crash files to a Geographic Information System (GIS)-based street network, identifies crash locations on the street network, and subsequently extracts information on road design and road geometry variables for each crash. As a result, for each year, the CRIS contains the characteristics of crashes occurring at intersection as well as non-intersection locations in Texas, along with supplementary information on road design and geometric variables.

For the current study, crashes at intersection locations were extracted out from the CRIS data base.[9] Further, we confined the analysis to intersections from the city of Arlington. This is

---

[7] In the broader crash study literature, a similar approach to Mitra has been adopted by Aguero-Valverde and Jovanis (2006, 2010), Miaou *et al.* (2003) and Song *et al.* (2006).

[8] The Texas law enforcement agency officially maintains the records of those crashes reported by police and drivers that involve property damage of more than $1,000 and/or the injury of one or more individuals (of course, records of crashes that involve fatalities on the spot are also maintained). Thus, the CRIS does not include minor crashes that involve only property damage of less than $1,000. However, in the rest of this paper, we will not belabor over this distinction, and will use the CRIS crashes as the measure for all crashes.

[9] TxDOT defines a crash as being intersection-related if it occurs within the curb-line limits of the intersection or on one of the approaches/exits to the intersection within 200 feet from the intersection center point.

because the CRIS does not include traffic flow information on intersection approach movements, one of the most relevant variables to explain intersection crash risk propensity (see Mountain *et al.*, 1998, Noland and Quddus, 2004, Quddus, 2008). So, we had to locate areas in the State that had good roadway segment traffic count data for the period from 2003-2009. Fortunately, the Arlington City Department of Public Works and Transportation maintains such traffic count data on Arlington streets. These traffic counts are two-way counts over a 24-hour period (excluding weekends). From this traffic count data, we extracted out the flows on the approach streets for each intersection in the Arlington traffic database. The sum of the flows on all approach streets to an intersection was computed to obtain an estimate of the total daily entering traffic at the intersection. Finally, the intersections from the Arlington data base were matched with the TxDOT crash files, incorporating the traffic volumes to the final database.

The count of all traffic crashes per year at each Arlington intersection in the TxDOT CRIS database constitutes the dependent variable of analysis. The final estimation sample includes 170 intersections, with crash counts available at each of these intersections for each year from 2003 to 2009 (a period of 7 years).[10] The sample of intersections produces 1190 (=170×7) intersection-year observations. The total number of crashes in the sample is 3,503, corresponding to an average of 2.94 crashes per year per intersection. The number of crashes per year at an intersection varies from a minimum of 0 to a maximum of 20. Figure 1 presents the distribution of crashes per year across all intersection-year combinations (this is the dependent variable in our spatial-temporal crash analysis). Not surprisingly, the most frequent count of number of crashes per year is zero, with 306 of the 1190 intersection-year observations (25.7% of the sample) having no crashes. This excess of zeros cannot be handled by a standard Poisson model, but is not a problem in our proposed framework because of the flexible specification of the thresholds. At the same time, the frequency distribution indicates a long right tail, which is also easily

---

[10] Note that our sample formation procedure includes only those Arlington intersections for which traffic count data is available from the Arlington database and at which at least one crash occurred over the 7-year period. While there are some intersections in the Arlington traffic count data base at which no crashes occurred during the 7-year period (as reflected in their absence in the CRIS database), we do not have readily available information on roadway design and control attributes for these intersections (precisely because they do not appear in the CRIS data base). Thus, we did not include such zero-crash (over the 7-year period) intersection locations. However, because of the 7-year length period of the CRIS data base, there were less than 30 intersections at which no crashes occurred at all out of the 200 intersections in the Arlington traffic count data base. Thus, the bias of our sample toward crash-prone intersections is not substantial. In any case, given that the dependent variable used in our analysis is crash count per year, there are several intersection-year combinations for which zero crashes are present in the sample, as we discuss later in this section. Thus zero crashes per year is well represented in the sample.

accommodated in the proposed count framework by setting $\alpha_k = \alpha_K$ for all $k > K$ (see Section 2.1). In the current empirical analysis, we set $K=9$ based on extensive testing with alternative values.

Table 1 presents the sample characteristics of the 170 intersections, including the (a) number of entering roads, (b) roadway alignment-- whether all approach streets are straight with no vertical grades, or whether one or more approach streets have a horizontal curvature or a vertical grade close to the intersection (this determination was made by TxDOT when compiling the roadway geometry information), (c) approach roadway type combination -- whether all approach roadways are city streets or at least one approach road is a non-city street, (d) type of traffic control, (e) total daily entering traffic, and (f) flow split imbalance in traffic volumes between the approaches (discussed in detail later). Table 1 indicates that, for a majority of intersections (about 72% of intersections), the number of entering roads is four. In addition, there are a sizeable number of intersections with three entering roads, both in a T-shape form as well as a Y-shape form. In terms of roadway alignment, more than 95% of intersections had straight approach streets with no vertical grade, while the remaining intersections had at least one approach with a horizontal curvature and/or a vertical grade. The approach roadways to a vast majority of the intersections are city streets, with less than 6% of intersections having one or more approach roads that are not city streets (of these intersections, 90% are intersections on US or State highway frontage roads). The traffic control type statistics in Table 1 indicate that nearly 80% of the intersections have one of the following three types of traffic control: regular signal light, yield sign control (a yield sign on one more approaches, but no other form of control), and stop sign control (a stop sign on one or more approaches). Intersections with flashing light (one or more approaches having a flashing red or yellow light) are also represented in the sample, as are intersections with no control but a center stripe/divider on one or more approaches. About 7.6% of the intersections do not have any traffic control (such as residential street intersections) or some minimal form of traffic control (such as turn marks and marked lanes). For ease, we will refer to such intersections as having no traffic control.

The statistics for the total daily entering traffic volume at intersections in Table 1 shows a large variation with a minimum of 2,866 vehicles and a maximum value of 193,178 vehicles. Indeed, the standard deviation of this attribute across intersections is almost equal to the mean. For each intersection, we also defined a major road as the one carrying the higher traffic volume,

and a minor road as the one carrying the lower traffic volume. For three-legged T-shaped intersections and four-legged intersections, the definitions of the major and minor roads are straightforward. For three-legged Y-shaped intersections, it so happened in our sample that each of these intersections had two of the three approaches with the same road name. This allowed the identification of a major road and a minor road just as in the earlier cases. For the small percentage of intersections with "more than four approaches", we manually determined a major road orientation and a minor road orientation based on intersection geometry. Next, a flow split imbalance (or FSIMB) factor between the volumes on the major and minor roadways was computed as follows:

$$FSIMB = \frac{V_1 - V_2}{V_1 + V_2},$$

where V_1 and V_2 correspond to the daily traffic volumes on the major and minor roadways, respectively ($V_1 \geq V_2$). Note that the FSIMB factor takes a value between zero (when there is no imbalance in flows on the approach roads) and one (when there is complete imbalance in the flows, theoretically obtained when there is zero flow on the minor road). The mean FSIMB statistic is 0.43, with a minimum value of zero and a maximum value of 0.97. The mean value of 0.43 indicates that, on average, the major road volume is 2.5 times the minor road volume at the sampled intersections.

The final row of Table 1 provides information on the distance between intersections. This is a key variable used to generate spatial dependency effects through the spatial weight matrix. To obtain the distance between intersections, the latitude and longitude coordinates (in degrees) of the center point of each intersection (as coded by TxDOT in the CRIS database) were first translated into *x-y* coordinates. Based on these coordinates, the Euclidean distance was computed for each pair of intersections. As can be observed from Table 1, the average distance between intersections in the city is 4.42 miles, with a minimum distance of 0.05 miles and a maximum distance of 11.81 miles (this maximum distance corresponds roughly to the length of the line that runs across the entire city of Arlington in the north-south direction). The distance between intersections was used as a measure of spatial proximity, and formed the basis to develop the spatial weight matrix. Several functional forms of distance were considered in the construction of the weight matrix including inverse distance and its higher orders, inverse of exponential distance, and an indicator for distance less than a threshold value.

4.3. Variable Specification and Model Formulation

Many different variable specifications, functional forms, and variable interactions were considered from the list of variables in Table 1 to determine the final model specification. The effects of the number of entering roads, roadway alignment, approach road type combination, and type of traffic control were introduced as categorical variables, while the total daily entering traffic volume and the FSIMB factor were introduced as continuous variables. For the categorical variables, the base category used was as follows: (a) four entering roads for the "number of entering roads", (b) straight approach streets with no vertical grade for "roadway alignment", (c) all approach roads are city streets for "approach road type combination", and (d) no traffic control for "type of traffic control". In addition to the categorical variables just discussed, we also included year-specific dummy variables to capture the generally reducing trend in accidents over time. To do so, we used the year 2003 as the base, and introduced dummy variables for each of the other years. For continuous variables (total daily entering traffic volume and FSIMB factor), we tested alternative functional forms that included linear and non-linear forms, and also dummy variables for different ranges. All these variables were considered both in the latent variable (with random coefficients) and threshold specification based on the findings of previous research and intuition. Further, various interactions of the continuous and the categorical variables were also considered whenever adequate observations were available to test such interaction effects, such as between traffic volume and type of traffic control, traffic volume and roadway alignment, and number of entering roads and type of traffic volume. But none of these interaction terms came out to be statistically significant. The final model was obtained based on statistical fit, intuitiveness, and parsimony considerations. Our final specification includes all the variables described before but "roadway alignment", which was not statistically significant.

Several different model formulations were estimated, but we present only three specific model formulations in the current paper to keep the discussion focused (for ease in presentation, we will use the terms "models" to refer to "model formulations" in the remainder of this section). The first model is the flexible count model cast in the generalized ordered-response probit (GORP) framework. In the notation of Section 3.2, this model corresponds to $\rho = 0$, $\delta = 0$, and all elements of $\overline{\Omega}$ being simultaneously zero. As discussed in Section 3.2, this model is more general than the traditional Poisson count model (by allowing exogenous variable effects in the

25

latent propensity variable as well as the thresholds, and having flexible thresholds), but ignores spatial, temporal, and unobserved heterogeneity effects. We will refer to this first model as the *flexible count model* (the flexible count model provided a far superior data fit, relative to a standard Poisson count model, in terms of the adjusted composite likelihood ratio index (ADCLRT) test of Section 3.3). The second model is the *flexible count model with temporal effects*. This model relaxes the constraint that $\rho = 0$ and all elements of $\overline{\mathbf{\Omega}}$ are zero. That is, it allows time-varying and time-invariant effects, as well as unobserved heterogeneity effects. However, it maintains the restriction that there are no spatial effects ($\delta = 0$). Further empirical investigation with this temporal dependency specification indicated that ρ, the term that generates time-varying correlation effects, was not statistically significantly different from zero in our empirical context. This implies that all the temporal dependency is caused by time-invariant (intersection-specific) generic crash propensity effects and time-invariant (intersection-specific) effects of exogenous variables. The final model is the *flexible count model with spatial and temporal effects*. This is the most general model, which we discuss in a little more detail below.

The spatial weights in the *flexible count model with spatial and temporal effects* may be generated in one of several ways, including inverse distance and its higher orders, inverse of exponential distance, and an indicator for distance less than a threshold value. In the case that the weight matrix is based on a continuous (and decaying) representation of distance, the analyst may also explore alternative distance bands to select the pairs of observations for inclusion in the composite marginal likelihood (CML) estimation. The optimal distance band may be set based on minimizing the trace of the variance-covariance matrix given by $tr[V_{CML}(\hat{\boldsymbol{\theta}})]$. In the current effort, we computed $tr[V_{CML}(\hat{\boldsymbol{\theta}})]$ for seven distance bands (2, 3, 4, 5, 6, 7, and 11.81 miles), the last one representing the case of including all the $QT(QT-1)/2$ possible intersection-year pairs in the CML function. Our results did not show substantial variations in the trace value for different distance bands (regardless of the specific continuous functional form used to represent the distance separation), though the best estimator efficiency was obtained at about 2 miles for all continuous distance representations. Further, we also constructed a spatial weight matrix based on a discrete distance indicator, which essentially allows spatial dependence among intersections located within "x" miles of each other. For this spatial weight matrix construction,

26

we included all pairings in the CML function within the "x" miles threshold, because selecting only a subset of intersection-year pairings (as used for the continuous distance representation for the weight matrix) is not valid in the discrete distance representation (there is no decaying effect of distance within the discrete distance threshold). At the end, the spatial weight matrix constructed based on the continuous distance representation in the form of the inverse of exponential distance provided the best results in terms of data fit (based on the CLIC statistic presented in Section 3.3), and all CML estimations were pursued with a distance band of 2 miles. The next section discusses the results of the following three models in more detail: (1) the flexible count or the FC model, (2) the flexible count model with temporal effects or the FCT model (as characterized by time-invariant effects), and (3) the flexible count model with temporal and spatial effects or the FCTS model.

4.4. Model Estimation Results

Table 2 presents the estimation results. We first discuss the effects of variables on the long-term crash propensity (Section 4.4.1), then the variable effects on the thresholds that affect the "instantaneous" translation of the propensity to whether or not a crash occurs at any given time (and, therefore, the count of crashes over a certain time period) (Section 4.4.2), next the temporal and spatial effects (Section 4.4.3), and finally the model fit comparisons (Section 4.4.4). Note that, for dummy exogenous variables, the category that does not appear in the table is the base category, as defined in Section 4.3.

A quick note here before proceeding further. The results of the three models in Table 2 are not directly comparable, since the scales of the error terms are different. Also, the second model accommodates spatial heteroscedasticity through the time-invariant random coefficient effects, while the third model accommodates spatial heteroscedasticity through both the random coefficients as well as the spatial lag formulation.

4.4.1. Long Term Crash Propensity

The constant term in the long term crash propensity is normalized to zero, as discussed in Section 3.1.1. However, the FCT and FCTS models accommodate intersection-specific unobserved heterogeneity effects due to unobserved factors (such as perhaps pedestrian walkway continuity characteristics or curb radius attributes or other roadway geometry features). The

standard deviations on the constant clearly identify the presence of time-stationary intersection-specific unobserved factors.

The other variables that significantly affect the long term crash propensity include number of entering roads, type of traffic control, traffic volume, the flow split imbalance (FSIMB) factor and the year-specific dummy variables. *Prima facie*, the mean values of parameter estimates are similar in sign in all the three models. The results indicate that intersections with three entering roads are less prone to crashes than four-legged intersections, probably because the former type of intersections presents "fewer vehicle conflict points" than the latter (Abdel-Aty and Wang, 2006). Intersections with more than four entering roads present the lowest crash propensity. This is reasonable because drivers are likely to be intrinsically cautious as they approach intersections that do not have the usual "four-entering roads" configuration and perhaps also because of conflict-reducing design safeguards at such atypical intersections.

The results on the traffic control type variables indicate a lower mean crash propensity when there is some kind of control (except flashing light control) relative to no control at all (notice the negative signs on the coefficients of the control variables except on the flashing light control). Intersections with regular signal lights present less long term crash propensity, on average, compared to other control types. The random coefficients on the yield sign control in the FCT (flexible count model with temporal effects) and FCTS (flexible count model with temporal and spatial effects) models are interesting. In both these models, the mean and standard deviation coefficients on the yield sign control variable suggest that, other things being controlled for, 88-89% of the intersections with a yield control are safer than intersections without any control, though a small percentage (11-12%) of intersections with a yield control can be more crash risk-prone than intersections without any control. But the results from the FCT and FCTS models are very different when the effects of the yield and regular signal light controls are compared. The FCT model suggests that about 45% of intersections with a yield control are safer from a long term crash risk propensity than observationally equivalent regular signal light-controlled intersections, while the FCTS model indicates that yield controls are pretty much always not as effective as signal-controlled intersections in reducing long term crash propensity (however, as we will see later, signal control also has an effect on the instantaneous translation of propensity into crash outcomes). Both the FCT and FCTS models indicate that 76 out of 100

yield-controlled intersections are safer than observationally equivalent stop-controlled intersections, perhaps because yield signs define the right-of-way at intersections quite clearly, and are characterized by lower traffic volumes and better sight distance than intersections controlled by stop signs. The results from all models also consistently suggest that intersections controlled by flashing lights (red or yellow) on one or more approaches are the most crash-prone, an observation also made by Poškienė and Sokolovskij (2008). This may be a reflection of confusion on the part of drivers regarding how to respond on seeing a flashing light and/or because flashing lights are usually installed at intersection locations that may have poor line of sight on approaches. Finally, intersections with center stripes or dividers help reduce long term risk propensity relative to intersections with no control, because such visual or physical barriers help in clearly delineating (and separating) the paths of conflicting traffic flow movements.

The total daily entering traffic volume variable, as defined in Section 4.2, was introduced in several ways, but the best data fit was obtained using a simple logarithmic transformation of the daily entering volume. The results show, on average, the expected positive relationship between total entering volume and long-term crash propensity. This is a direct consequence of higher volumes being related to tighter vehicle spacing and more conflict points on a per time unit basis (see, for example, Chin and Quddus, 2003, Mitra, 2009, Oh *et al.*, 2009, and Abdel-Aty and Wang, 2006). The logarithmic functional form shows the marginally reducing effect of traffic volume on count frequency. However, unlike most other studies, we also are able to consider random coefficients in a way that is simpler and different from the usual mixing approach used in traditional count models (see Anastasopoulos and Mannering, 2009 and the discussion in Section 2.1). Our results show the clear heterogeneity in the influence of entering volume on long term crash propensity. In particular, the mean and standard deviation on the logarithm of the total daily entering volume variable in Table 2 reveal that an increase in the total daily entering traffic volume leads to a lower crash propensity at about a third of the intersections, but leads to a higher crash propensity at two-thirds of the intersections (this is for both the FCT and the FCTS models). Such variations may result from the complex interactions between unobserved intersection characteristics and motorist learning/adaptation behavior in response to different levels of traffic volume.

In this study, we introduce the concept of the Flow Split Imbalance (or FSIMB) factor to capture the difference in traffic volumes between the major and minor roadways. Such volume

differences have been considered in earlier safety studies too (see, for example, Wang *et al*., 2006, Haque *et al*., 2010, Huang and Chin, 2010), but not in the specific form we propose. The negative parameter in Table 2 associated with this FSIMB variable suggests a lower crash propensity associated with higher flow imbalance. That is, intersections where the volumes on the minor and major roadways are relatively unbalanced are less crash-prone than intersections where the minor and major roadways have about the same traffic volume. This perhaps reflects the caution that drivers exercise when approaching an intersection on a minor roadway when aware of the much higher traffic volume on the other roadway. Besides, the dominant flow on one of the roadways should reduce the number of conflict points.

Finally, the year-specific dummy variables show the lower crash propensity in recent years relative to earlier years, which is consistent with the reduction in crashes in recent years (see, for example, Aguero-Valverde and Jovanis, 2006 and Quddus, 2008).

4.4.2 Threshold Parameters

The threshold parameters include the threshold specific constants (α_k values), as well as a constant and variables associated with approach roadway type combination and type of traffic control as part of γ vector (see Equation (8)). The thresholds are responsible for the "instantaneous" translation of the long-term crash propensity to whether or not a crash occurs at any given time (that is, they determine the mapping of the latent propensity to the observed count outcome).

The threshold specific constants (α_k) do not have any substantive interpretations. However, their presence provides flexibility in the count model to accommodate high or low probability masses for specific outcomes. As indicated in Section 2.1, identification is achieved by specifying $\alpha_0 = 0$ and $\alpha_k = \alpha_K \ \forall \ k \geq K$. In the present specification, we initially set $K = 20$ and progressively reduced K based on statistical significance considerations and general data fit. The final specifications in Table 2 are based on setting $K = 9$.

The elements in the γ vector are presented next in Table 2. The constant does not have any particular interpretation. For the other variables, a positive coefficient shifts all the thresholds toward the left of the crash propensity scale, which has the effect of reducing the probability of zero crashes. The effect of the approach roadway type combination variable

suggests that, given two intersections with the same long term crash propensity, the intersection where at least one approach roadway is a non-city street (such as a highway or interstate frontage road) is more likely to have a non-zero crash outcome compared to the intersection where all approaching roads are city streets. This is a result that certainly warrants in-depth analysis, but may be suggestive of motorists not reducing speed enough after exiting off a highway as they approach an intersection on a frontage road. The other effect of regular signal light as the traffic control is interesting, and indicates an increase in non-zero crash outcomes at intersections with regular signal lights relative to other types of control (for a given long-term crash propensity). That is, the translation of risk into the occurrence of a crash is elevated for regular signal light controlled intersections, perhaps because of less of an "out" option at signal controlled intersection as a crash is developing (for example, motorists may not be able to get into a different lane or maneuver in a different direction at the intersection because of other simultaneous movements taking place and because of the clearly delineated and channeled traffic movements). The suggestion is that movement delineation and separation can be a double-edged sword – while it reduces conflict points and risks of a crash, it also provides fewer options to motorists to avoid a crash situation as it starts to develop.

4.4.3. Temporal and Spatial Effects

As discussed earlier, the unique feature of our formulation is that it enables the accommodation of temporal and spatial effects through the long-term propensity variable in the generalized ordered-response characterization of a count model. The statistically significant estimates of the random coefficients on the constant, on the indicator variable for yield control, and on the logarithm of daily traffic counts reflect the presence of time-invariant intersection-specific temporal dependence effects on the long term crash propensity. However, the first order auto-regressive temporal dependency parameter ρ did not turn out to be statistically significant.

The spatial dependency parameter δ in the final spatial lag model is positive, moderate in magnitude (about 0.422), and statistically significant, supporting the hypothesis that factors such as intersection geometric configuration, traffic control type, and traffic volumes will have "spillover" effects at other intersections. Also, the spatial lag model reinforces the notion of the presence of unobserved factors (such as zonal regulations, neighborhood design features, and

neighborhood driving attitudes) that generate correlation in crash occurrence at proximately located intersections.

4.4.4. Model Selection and Statistical Fit

The flexible count model with temporal effects (FCT) is superior to the flexible count (FC) model, as should be clear from the presence of statistically significant random coefficients. Further, the FCTS model shows a statistically significant spatial lag effect. Another way to demonstrate these improvements is by undertaking the adjusted composite likelihood ratio test (ADCLRT) test. The composite log-likelihood value for the FC model is -468,491.1 (27 parameters estimated), for the FCT model is -455,532.0 (30 parameters estimated), and for the FCTS model (31 parameters estimated) is -455,195.2. The ADCLRT statistic for the test between the FCTS and FC models is 1646, which is larger than the table chi-squared value with 4 degrees of freedom at any reasonable level of significance. Similarly, the ADCLRT statistic for testing the FCTS model with the FCT model turned out to be 356, which is again higher than the chi-squared table value with one degree of freedom at any reasonable level of significance.

4.5. Elasticity Effects

In the previous section we concluded that the FCTS model is statistically superior to the other models presented in Table 2. However, and very importantly, the difference between these models is not simply a matter of data fit. We expect that the impact of variables on crash frequency will be different among the models. The parameters of the exogenous factors in Table 2, however, do not directly provide the magnitude of the impact of variables on crash frequency. To do so, we compute the aggregate-level "elasticity effects" of variables to discern the magnitude and direction of variable impacts. Specifically, we examine the effects of variables on the expected number of crashes at any intersection for the year 2009, given the intersection characteristics and those of other intersections. The expected number of crashes at intersection q is computed as:

$$E(y_{q,2009}) = \sum_{k=0}^{\infty} P(y_{q,2009} = k) \cdot k,$$

(19)

where $P(y_{q,2009} = k)$ is the probability of k crashes occurring at intersection q for the year 2009. Although the summation in the equation above extends until infinity, we consider counts only up to $k = 20$, which is the maximum crash frequency observed in the dataset. This should not affect the elasticity computations because the probabilities associated with higher crash outcomes are very close to zero. Also, using the notation in Section 3.1, the expected value of crashes at intersection q is a function of variables in the $(Q \times L)$ matrix of exogenous variables for all Q units $x_{2009} = (x_{1,2009}, x_{2,2009}, x_{3,2009},..., x_{Q,2009})'$ as well as a function of variables in the $z_{q,2009}$ vector embedded in the thresholds:

$$\psi_{q,k,2009} = \Phi^{-1}\left(e^{-\lambda_{q,2009}} \sum_{l=0}^{k} \frac{\lambda_{q,2009}^{l}}{l!} \right) + \alpha_k, \ \lambda_{q,2009} = e^{\gamma' z_{q,2009}}, \ \alpha_0 = 0, \ \alpha_k = \alpha_9 \text{ if } k \geq 9. \tag{20}$$

If there are common variables in $x_{q,2009}$ and $z_{q,2009}$ (such as the "regular signal light" variable in the current application), these variables will impact the expected value of the crash frequency at intersection q both through the long-term propensity and the thresholds.

The estimate of $P(y_{q,2009} = k)$ in Equation (19) for the FC model may be obtained in a straightforward manner from Equation (2). For the FCT model, random coefficients need to be accommodated and, for the FCTS model, spatial effects also need to be recognized. For ease, we will focus on the FCTS model, since the procedure to estimate $P(y_q = k)$ for the FCT model is simpler. For the FCTS model, using the same notations as in Section 3.1, we may write the following counterpart of Equation (10) for the year 2009:

$$y_{2009}^* = \widetilde{S} x_{2009} b + \widetilde{S} \widetilde{x}_{2009} \widetilde{\beta} + \widetilde{S} \varepsilon_{2009}, \text{ where } \widetilde{S} = \left[(I_Q - \delta W)^{-1} \right]. \tag{21}$$

To estimate $P(y_{q,2009} = k)$ in the FCTS model, we simulate the above $Q \times 1$-vector y_{2009}^* thousand times using the estimated values of δ, b, and draw randomly 1000 times from the appropriate normal distributions for $\widetilde{\beta}$ and the $Q \times 1$-vector ε_{2009}. Next, we compare each of the 1000 draws of the qth element of y_{2009}^* with the corresponding thresholds for the qth element from Equation (20), assign the count value for each of the 1000 draws based on this comparison, and then take the share of each count prediction across the 1000 draws to estimate $P(y_{q,2009} = k)$. We also

compute the standard errors of the elasticity effects by using 200 bootstrap draws from the sampling distributions of the estimated parameters.[11]

The elasticity computed is a measure of the aggregate percentage change in the crash frequency due to a change in an exogenous variable. For dummy variables, the value of the variable is changed to one for the subsample of intersections for which the variable takes a value of zero, and to zero for the subsample of observations for which the variable takes a value of one. We then add the shifts in expected aggregate shares in the two subsamples after reversing the sign of the shifts in the second subsample, and compute the effective percentage change in the expected total number of crashes per year across all intersections in the sample due to a change in the dummy variable from 0 to 1. For continuous variables, we increase the value of the variable by 10% for each intersection and compute the percentage change in the expected total number of crashes per year across all intersections. For the FSIMB factor that is contained between 0 and 1, we increase the factor by 0.1 at each intersection.

The elasticity effects and their standard errors are computed for the FC model and the FCTS model, and are presented in Table 3 (we focus only on the FC and FCTS models in this section to keep the presentation concise). The first entry in the table indicates that the number of crashes at three-legged intersections is, on average, about 51% less than the number of crashes at four-legged intersections. Other entries may be similarly interpreted. The elasticity effects of both the FC and FCTS models are in the same direction, and are consistent with the discussions in the previous section. As should be clear from the table, the magnitude of effects is, in general, higher in the FCTS model than in the FC model. This is because of the "spillover" effects in the FCTS model, which causes a spatial multiplier effect. A change in a variable at an intersection A affects the frequency of crashes at other intersections that then comes back and has an impact on the crashes at intersection A. The FC model does not capture such spatial multiplier effects because it considers crashes at one intersection to be independent of crashes at other intersections. The only exception to the spatial multiplier effect relates to the variable "at least one approach roadway is a non-city street". The magnitude of effect for this variable is higher in the FC model relative to the FCTS model, but this is because the variable affects crash frequency only through the thresholds and not through the long-term crash propensity variable involved in

---

[11] For ease in computation, we fix the spatial lag parameter δ in the bootstrapping, so that we do not have to compute the matrix $\widetilde{S}$ for each bootstrap draw (the matrix $\widetilde{S}$ entails a high-dimensional matrix inversion).

the spillover effects. The elasticity results of the other variables in Table 3 indicate a lower reduction in crash frequency as implied by the FC model for intersections with more than four entering roads (relative to four entering roads) and a much lower impact of the logarithm of daily entering volume on crash occurrence at intersections (these differences in predictions between the models are statistically significant at the 0.02 level of significance). At the same time, relative to the FCTS model, the FC model predicts a much higher crash frequency rate at intersections where at least one approach roadway is a non-city street (mostly intersections on US or State highway frontage roads) compared to usual urban intersections (this difference in predictions is statistically significant at the 0.06 level of significance). The FC model also under-estimates the reduction in crashes at intersections with higher flow-split imbalance (*i.e.*, more flow disparity in approaches) relative to the FCTS model (statistically significant difference at the 0.12 level of significance), and predicts a lower crash rate at intersections with flashing lights compared to the FCTS model (statistically significant at about the 0.2 level of significance). Overall, there are statistically significant differences in elasticity predictions between the models, highlighting the potentially misinformed investments in crash countermeasures if temporal and spatial effects are ignored.

Across both the FC and FCTS models, the two most important determinants of high crash frequency are flashing light control and non-city street approaches. The flashing light control effect may simply be proxying for various elements of the intersection location that warrant a flashing light installation in the first place (such as the installation of flashing lights to catch the attention of motorists on approaches with poor line of sight, or regions of fast moving traffic conflicting with rural side streets). But it may also indicate confusion in the minds of motorists regarding the purpose and intent of flashing lights. If not clear, motorists can become insensitive to such installations and ignore them, or may not know how to respond on seeing such lights. The high magnitude of effect of non-city street approaches suggests, as already indicated earlier, that motorists may not be reducing speed enough after exiting off a highway as they approach an intersection on a frontage road. Further investigation of these effects will be helpful to understand the precise reasons for this result, which can in turn lead to improved intersection designs as well as appropriate outreach and dissemination campaigns to inform the driving public.

5. CONCLUSIONS

In the current paper, we propose a count modeling framework and inference approach that resolves many challenges in extant models for correlated counts. Specifically, we show how any traditional count model can be reformulated as a special case of a generalized ordered response model in which a single latent continuous variable is partitioned into mutually exclusive intervals. Using this equivalent latent variable-based generalized ordered response framework for count data models, we are then able to gainfully introduce spatial dependencies using a spatial structure on the latent continuous variables, and time-stationary and time-varying temporal correlation patterns by means of an appropriate structure for the error term of the latent variable. Our formulation also allows handling excess zeros in correlated count data, a phenomenon that is commonly found in practice. A composite marginal likelihood inference approach for ordered-response models is used to estimate model parameters. This approach reduces the inference problem dimension to pairs of bivariate probability computations, obtaining consistent and asymptotically normally distributed estimates. The approach is easy to implement using available discrete choice software programs or matrix programming languages.

The modeling framework is applied to predict crash frequency at urban intersections in Arlington, Texas. There have been several efforts devoted to investigating crash occurrence; however, most of these studies ignore the presence of temporal correlation across repeated data from the same intersection and spatial dependence across intersections. Temporal correlations can manifest themselves in the form of random intersection-specific coefficients on variables impacting crash propensity at the intersection, as well as in the form of time-varying correlation effects. Spatial correlation can emerge when observed factors have a "spillover" effect on crash frequency, or when spatially unobserved factors generate spatial correlations in crash frequencies at closely located intersections. Ignoring such temporal and spatial correlations will, in general, lead to inconsistent and inefficient parameter estimates.

The empirical results reveal the presence of intersection-specific time-invariant unobserved components influencing crash propensity and a spatial lag structure to characterize spatial dependence. Roadway configuration, approach roadway functional types, traffic control type, total flows and the split of flows between approaches are all important variables in determining crash frequency at intersections. The results highlight the potentially misinformed effects of these variables on crash frequency that can result if temporal dependencies and spatial

36

dynamics are ignored. Future crash analysis using the method proposed here may be undertaken with data sets that provide additional geometric design characteristics of intersections (such as number of lanes on approaches, presence of separate left-turning lanes, speed limits, and permitted movements) and land use attributes at the intersection location.

The method proposed here is quite general, and may be used to analyze any number of correlated count outcomes with relative ease. For instance, the empirical analysis in the current paper may be extended to model crash counts by severity level, while accommodating global and flexible spatial and temporal interactions.

REFERENCES

Abdel-Aty, M., Wang, X., 2006. Crash estimation at signalized intersections along corridors: analyzing spatial effect and identifying significant factors. *Transportation Research Record* 1953, 98-111.

Aguero-Valverde, J., Jovanis, P., 2006. Spatial analysis of fatal and injury crashes in Pennsylvania. *Accident Analysis and Prevention* 38(3), 618-625.

Aguero-Valverde, J., Jovanis, P., 2010. Spatial correlation in multilevel crash frequency models effects of different neighboring structures. *Transportation Research Record* 2165, 21-32.

Anastasopoulos, P., Mannering, F., 2009. A note on modeling vehicle accident frequencies with random-parameters count models. *Accident Analysis and Prevention* 41(1), 153-159.

Apanasovich, T.V., Ruppert, D., Lupton, J.R., Popovic, N., Turner, N.D., Chapkin, R.S., Carroll, R.J., 2008. Aberrant crypt foci and semiparametric modelling of correlated binary data. *Biometrics* 64(2), 490-500.

Bermúdez, L., Karlis, L., 2011. Bayesian multivariate Poisson models for insurance ratemaking. *Insurance: Mathematics and Economics* 48(2), 226-236.

Beron, K.J., Vijverberg, W.P.M., 2004. Probit in a spatial context: A Monte Carlo analysis. In *Advances in Spatial Econometrics: Methodology, Tools and Applications*, Anselin, L., Florax, R.J.G.M., Rey, S.J. (eds.), Springer-Verlag, Berlin, 169-196.

Beron, K.J., Murdoch, J.C., Vijverberg, W.P.M., 2003. Why cooperate? Public goods, economic power, and the Montreal protocol. *Review of Economics and Statistics* 85(2), 286-297.

Bethell, J., Rhodes, A. E., Bondy, S. J., Lou, W. Y. W., Guttmann, A., 2010. Repeat self-harm: application of hurdle models. *The British Journal of Psychiatry* 196(3), 243-244.

Bhat, C.R., 2011. The maximum approximate composite marginal likelihood (MACML) estimation of multinomial probit-based unordered response choice models. *Transportation Research Part B* 45(7), 923-939.

Bhat, C.R., Sener, I.N., 2009. A copula-based closed-form binary logit choice model for accommodating spatial correlation across observational units. *Journal of Geographical Systems* 11(3), 243-272.

Bhat, C.R., Varin, C., Ferdous, N., 2010a. A comparison of the maximum simulated likelihood and composite marginal likelihood estimation approaches in the context of the multivariate ordered response model. In *Advances in Econometrics: Maximum Simulated Likelihood Methods and Applications*, Greene, W., Hill, R.C. (eds.), Emerald Group Publishing Limited, 65-106.

Bhat, C.R., Sener, I.N., Eluru, N., 2010b. A flexible spatially dependent discrete choice model: formulation and application to teenagers' weekday recreational activity participation. *Transportation Research Part B* 44(8-9), 903-921.

Buck, A., Blackstone, E., Hakim, S., 2009. A multivariate poisson model of consumer choice in a multi-airport region. *iBusiness* 1(2), 85-98.

Caragea, P.C., Smith, R.L., 2007. Asymptotic properties of computationally efficient alternative estimators for a class of multivariate normal models. *Journal of Multivariate Analysis* 98(7), 1417-1440.

Chin, H., Quddus, M., 2003. Applying the random effect negative binomial model to examine traffic accident occurrence at signalized intersections. *Accident Analysis and Prevention* 35(2), 253-259.

Cox, D.R., Reid, N., 2004. A note on pseudolikelihood constructed from marginal densities. *Biometrika* 91(3), 729-737.

Eluru, N., Bhat, C.R., Hensher, D.A., 2008. A mixed generalized ordered response model for examining pedestrian and bicyclist injury severity level in traffic crashes. *Accident Analysis and Prevention* 40(3), 1033-1054.

Engle, R.F., Shephard, N., Sheppard, K., 2007. Fitting and testing vast dimensional time-varying covariance models. Finance Working Papers, FIN-07-046, Stern School of Business, New York University.

Engler, D.A., Mohapatra, G., Louis, D.N., Betensky, R.A., 2006. A pseudolikelihood approach for simultaneous analysis of array comparative genomic hybridizations. *Biostatistics* 7(3), 399-421.

Fleming, M.M., 2004. Techniques for estimating spatially dependent discrete choice models. In *Advances in Spatial Econometrics: Methodology, Tools and Applications*, Anselin, L., Florax, R.J.G.M., Rey, S.J. (eds.), Springer-Verlag, Berlin, 145-168.

Franzese, R.J., Hays, J.C., 2008. Empirical models of spatial interdependence. In *The Oxford Handbook of Political Methodology*, Box-Steffensmeier, J.M., Brady, H.E., Collier, D., (eds.), Oxford University Press, Oxford, 570-604.

Franzese, R.J., Hays, J.C., Schaffer, L., 2010. Spatial, temporal, and spatiotemporal autoregressive probit models of binary outcomes: estimation, interpretation, and presentation. APSA 2010 Annual Meeting Paper. Available at SSRN: http://ssrn.com/abstract=1643867

Godambe, V.P., 1960. An optimum property of regular maximum likelihood estimation. *The Annals of Mathematical Statistics* 31(4), 1208-1211.

Greene, W.H., 2005. Censored data and truncated distributions. In Mills, T.C., Patterson, K. (eds.), *Palgrave Handbook of Econometrics, Volume 1: Econometric Theory*, Palgrave Macmillan, Hampshire.

Greene, W.H., Hensher, D.A., 2010. *Modeling Ordered Choices: A Primer*. Cambridge University Press, Cambridge.

Griebe, P., 2003. Accident prediction models for urban roads. *Accident Analysis and Prevention* 35(2), 273-285

Haque, M., Chin, H., Huang, H., 2010. Applying Bayesian hierarchical models to examine motorcycle crashes at signalized intersections. *Accident Analysis and Prevention* 42(1), 203-212.

Heagerty, P., Lumley, T., 2000. Window subsampling of estimating functions with application to regression models. *Journal of the American Statistical Association* 95(449), 197-211.

Herriges, J.A., Phaneuf, D.J., Tobias, J.L., 2008. Estimating demand systems when outcomes are correlated counts. *Journal of Econometrics* 147(2), 282-298.

Huang, H., Chin, H., 2010. Modeling road traffic crashes with zero-inflation and site-specific random effects. *Statistical Methods and Applications* 19(3), 445-462.

Joe, H., Lee, Y., 2009. On weighting of bivariate margins in pairwise likelihood. *Journal of Multivariate Analysis* 100(4), 670-685.

Kapteyn, A., Smith, J.P., van Soest, A., 2007. Vignettes and self-reports of work disability in the United States and the Netherlands. *American Economics Review* 97(1), 461-473.

King, G., 2009. The Anchoring Vignettes Website. At http://gking.harvard.edu/vign.

King, G., Murray, C., Salomon, J., Tandon, A., 2004. Enhancing the validity and cross-cultural comparability of measurement in survey research. *American Political Science Review* 98(1), 191-207.

Kuk, A., Nott, D., 2000. A pairwise likelihood approach to analyzing correlated binary data. *Statistics and Probability Letters* 47(4), 329-335.

LeSage, J. P., Pace R., 2009. *Introduction to Spatial Econometrics*. CRC Press Taylor & Francis Group, Boca Raton, London, New York.

Lindsay, B.G., 1988. Composite likelihood methods. *Contemporary Mathematics* 80, 221-239.

Malyshkina, N.V., Mannering, F.L., 2010. Zero state Markov switching count-data models: An empirical assessment, *Accident Analysis and Prevention*, 42(1), 122-130.

Miaou, S., Song, J., Mallick, B., 2003. Roadway traffic crash mapping: a space–time modeling approach. *Journal of Transportation and Statistics* 6(1), 33-57.

Mitra, S., 2009. Spatial autocorrelation and Bayesian spatial statistical method for analyzing intersections prone to injury crashes. *Transportation Research Record* 2136, 92-100.

Molenberghs, G., Verbeke, G., 2005. *Models for Discrete Longitudinal Data*. Springer Series in Statistics, Springer Science + Business Media, Inc., New York.

Mountain, L., Maher, M., Fawaz, B., 1998. The influence of trend on estimates of accidents at junctions. *Accident Analysis and Prevention* 30(5), 641-649.

Müller, G., Czado, C., 2005. An autoregressive ordered probit model with application to high frequency financial data. *Journal of Computational and Graphical Statistics* 14(2), 320-338.

Musio, M., Sauleau, E., Buemi, A., 2010. Bayesian semi-parametric ZIP models with space–time interactions: an application to cancer registry data. *Mathematical Medicine and Biology* 27(2), 181-194.

National Highway Traffic Safety Administration (NHTSA), 2010a. Traffic Safety Fact Sheet, 2009. Publication DOT HS 811392, National Center for Statistics and Analysis, National Highway Traffic Safety Administration.
Available at: http://www-nrd.nhtsa.dot.gov/Pubs/811392.pdf

National Highway Traffic Safety Administration (NHTSA), 2010b. Crash Factors in Intersection-Related Crashes: An On-Scene Perspective. Publication DOT HS 811366, National Center for Statistics and Analysis, National Highway Traffic Safety Administration.
Available at: http://www-nrd.nhtsa.dot.gov/Pubs/811366.pdf

Noland, R., Quddus, M., 2004. A spatially disaggregate analysis of road casualties in England. *Accident Analysis and Prevention* 36(6), 973-984.

Oh, J., Washington, S., Lee, D., 2009. Expected safety performance of rural signalized intersections in South Korea. *Transportation Research Record* 2114, 72-82.

Pace, L., Salvan, A., Sartori, N., 2011. Adjusting composite likelihood ratio statistics. *Statistica Sinica* 21(1), 129-148.

Phaneuf, D.J., Palmquist, R.B., 2003. Estimating spatially and temporally explicit land conversion models using discrete duration.
Unpublished: http://www.aere.org/meetings/0306workshop_Phaneuf.pdf.

Poškienė, K., Sokolovskij, E., 2008. Traffic control elements influence on accidents, mobility and the environment. *Transport* 23(1), 55-58.

Pudney, S.E., Shields, M.A., 2000. Gender, race, pay and promotion in the British nursing profession: estimation of a generalised ordered probit model. *Journal of Applied Econometrics* 15(4), 367-399.

Quddus, M., 2008. Time series count data models: an empirical application to traffic accidents. *Accident Analysis and Prevention* 40(5), 1732-1741.

Smith, T., LeSage, J.P., 2004. A Bayesian probit model with spatial dependencies. In *Advances in Econometrics: Volume 18: Spatial and Spatiotemporal Econometrics*, LeSage, J.P., Pace, R.K. (eds.), Elsevier Ltd, Oxford, 127-160.

Song, J.J., Ghosh, M., Miaou, S., Mallick, B., 2006. Bayesian multivariate spatial models for roadway traffic crash mapping. *Journal of Multivariate Analysis* 97(1), 246-273.

Terza, J.V., 1985. Ordinal probit: a generalization. *Communication in Statistics Theory and Methods* 14(1), 1-11.

U.S. National Safety Council, 2009. Estimating the costs of unintentional injuries. Available at: http://www.nsc.org/news_resources/injury_and_death_statistics/Pages/EstimatingtheCostsof UnintentionalInjuries.aspx

Varin, C., Czado, C., 2010. A mixed autoregressive probit model for ordinal longitudinal data. *Biostatistics* 11(1), 127-138.

Varin, C., Vidoni, P., 2009. Pairwise likelihood inference for general state space models. *Econometric Reviews* 28(1-3), 170-185.

Varin, C., Reid, N., Firth, D., 2011. An overview of composite likelihood methods. *Statistica Sinica* 21(1), 5-42.

Ver Hoef, M., Jansen, J., 2007. Space–time zero-inflated count models of harbor seals. *Environmetrics* 18(7), 697-712.

Wang, X., Abdel-Aty, M., 2006. Temporal and spatial analysis of rear-end crashes at signalized intersections. *Accident Analysis and Prevention* 38(6), 1137-1150.

Wang, X., Abdel-Aty, M., Brady, P., 2006. Crash estimation at signalized intersections significant factors and temporal effect. *Transportation Research Record* 1953, 10-20.

Winkelmann, R., 2000. *Econometric Analysis of Count Data*. Springer-Verlag, New York.

Xu, X., Reid, N., 2011. On the robustness of the maximum composite likelihood estimate, *Journal of Statistical Planning and Inference* 141(9), 3047-3054.

Yi, G.Y., Zeng, L., Cook, R.J., 2011. A robust pairwise likelihood method for incomplete longitudinal binary data arising in clusters. *The Canadian Journal of Statistics*, 39(1), 34-51.

Zavoina, R., McKelvey, W., 1975. A statistical model for the analysis of ordinal-level dependent variables. *Journal of Mathematical Sociology* 4(1), 103-120.

Zhao, Y., Joe, H., 2005. Composite likelihood estimation in multivariate data analysis. *The Canadian Journal of Statistics* 33(3), 335-356.

LIST OF FIGURES
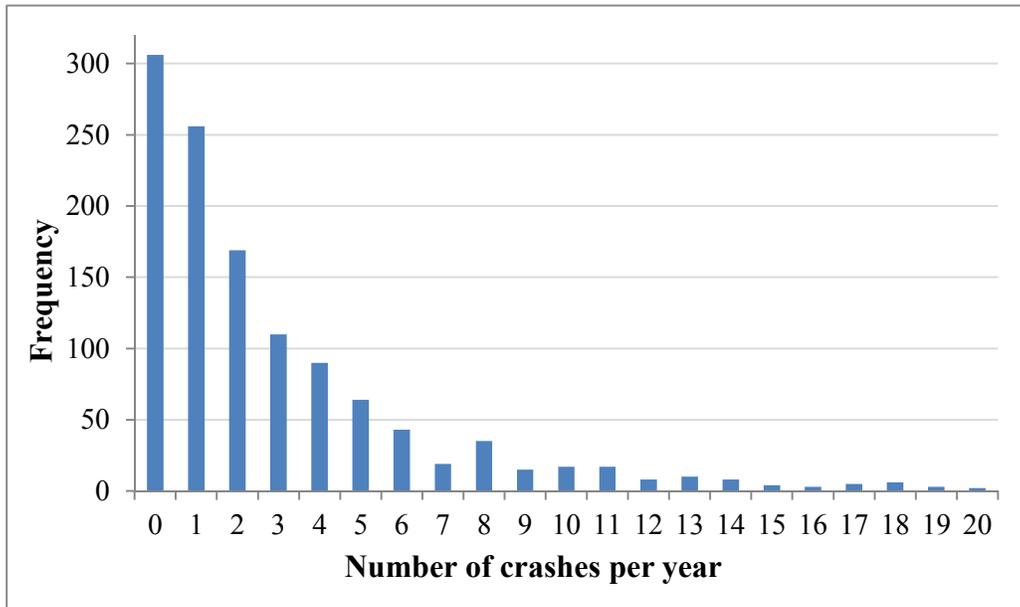
LIST OF TABLES

Figure 1. Yearly Crash Frequency Distribution Across Intersections

Table 1. Sample Characteristics (1190 observations)

Variables	Sample share
Number of Entering Roads	
Three	24.6
Four	71.6
More than four	3.8
Roadway Alignment	
All approaches are straight with no vertical grade	95.7
At least one approach has horizontal curvature or vertical grade	4.3
Approach Roadway Type Combination	
All approach roadways are city streets	94.4
At least one approach roadway is a non-city street	5.6
Type of Traffic Control	
Regular signal light	52.8
Yield sign	15.3
Stop sign	12.4
Flashing light	7.3
Center stripe / divider	4.6
No traffic control or minimal traffic control	7.6

Descriptive Statistics

	Minimum	Maximum	Mean	Std. Dev.
Total daily entering volume (vehicles/day)	2,866	193,178	35,222	33,784
Flow split imbalance (FSIMB) factor	0.00	0.97	0.43	0.25
Distance between intersections (miles)	0.05	11.81	4.42	2.29

Table 2. Model Estimation Results
(Weight Matrix: inverse of exponential distance, Distance Band: 2 miles)

Variables	Flexible Count (FC) Model		Flexible Count with Temporal Effects (FCT) Model		Flexible Count with Temporal and Spatial Effects (FCTS) Model	
	Estimate	t-stat	Estimate	t-stat	Estimate	t-stat
Long-term Propensity Variables						
Constant	0.000	-	0.000	-	0.000	-
Standard Deviation	-	-	0.889	10.33	0.640	7.02
Number of Entering Roads						
Three	-0.660	-15.07	-1.048	-12.95	-0.950	-13.20
More than four	-0.921	-13.64	-1.397	-14.61	-1.276	-13.20
Type of Traffic Control						
Regular signal light	-1.906	-7.35	-1.709	-3.78	-3.003	-7.53
Yield sign	-0.912	-10.40	-1.536	-8.32	-1.352	-8.55
Standard Deviation	-	-	1.279	6.04	1.168	6.32
Stop sign	-0.454	-7.80	-0.630	-6.93	-0.540	-7.01
Flashing light	0.696	8.69	1.002	8.99	0.948	9.08
Center stripe / divider	-0.659	-8.83	-0.965	-7.64	-0.786	-6.96
Logarithm of daily entering volume (veh/day/10,000)	0.210	6.76	0.437	8.42	0.374	7.98
Standard Deviation	-	-	0.966	16.66	0.831	15.60
Flow split imbalance (FSIMB) factor	-0.817	-9.86	-1.146	-9.39	-1.042	-9.49
Year-specific Dummy Variables						
Year 2004	-0.210	-2.23	-0.267	-1.46	-0.103	-1.09
Year 2005	-0.440	-4.67	-0.652	-3.63	-0.291	-2.90
Year 2006	-0.438	-4.69	-0.682	-3.74	-0.218	-1.94
Year 2007	-0.441	-4.67	-0.690	-3.76	-0.182	-1.52
Year 2008	-0.470	-5.03	-0.748	-4.10	-0.250	-2.10
Year 2009	-0.539	-5.85	-0.835	-4.62	-0.378	-3.30
Threshold Variables						
Threshold Specific Constants						
α_1	-0.026	-1.06	0.519	9.04	0.492	10.30
α_2	-0.166	-5.02	0.740	8.76	0.708	10.16
α_3	-0.377	-8.81	0.787	7.37	0.769	8.77
α_4	-0.563	-11.25	0.856	6.69	0.842	7.93
α_5	-0.745	-13.02	0.913	6.12	0.901	7.23
α_6	-0.968	-15.18	0.885	5.29	0.882	6.20
α_7	-1.264	-17.25	0.698	3.79	0.722	4.65
α_8	-1.430	-18.88	0.792	3.82	0.809	4.65
α_9	-1.698	-20.81	0.648	2.93	0.689	3.70
γ *Vector*						
Constant	1.223	40.40	1.741	19.56	2.309	16.09
Approach Roadway Type Combination						
At least one approach roadway is a non-city street	0.449	11.39	0.418	9.37	0.313	9.00
Type of Traffic Control						
Regular signal light	0.837	7.49	0.622	3.89	0.888	10.82
δ (spatial correlation parameter)	-	-	-	-	0.422	9.25
Number of observations	1190		1190		1190	
Number of parameters estimated	27		30		31	
Log-composite likelihood at convergence	-468,491.1		-455,532.0		-455,195.2	

45

Table 3. Elasticity Effects of Variables on Expected Number of Crashes

Variable	Flexible Count (FC) Model		Flexible Count with Spatial and Temporal Effects (FCTS) Model	
	Estimate	Standard error	Estimate	Standard error
Number of Entering Roads				
Three	-51.07	4.31	-54.44	5.63
More than four	-60.95	4.07	-75.08	4.39
Approach Roadway Type Combination				
At least one approach roadway is a non-city street	133.44	20.20	85.22	15.63
Type of Traffic Control				
Regular signal light	-2.85	33.24	-7.87	36.31
Yield sign	-59.45	5.04	-69.50	6.78
Stop sign	-34.79	4.78	-37.50	5.83
Flashing light	69.45	11.94	99.40	21.43
Center stripe / divider	-47.36	5.53	-53.71	7.42
Logarithm of daily entering volume (veh/day/10,000) – 10% increase	1.75	0.31	5.06	1.09
Flow split imbalance (FSIMB) factor – 0.1 increase	-6.88	0.98	-9.74	1.65